

Aptitudes des communautés de recherche en ORD

Mandat 2 mandaté par Swissuniversities

Réalisé par la HEG, Mars - Octobre 2021
Rapport V2 - 22/10/2021

Auteurs :

Gaia Bonghi
Marielle Guirlet
Elise Pelletier
Grégoire Urvoy
René Schneider

Table des matières

Table des figures.....	4
Table des tableaux.....	4
Sigles et acronymes.....	5
Résumé.....	6
Introduction et contexte	7
Méthodologie.....	8
1) Objectifs.....	8
2) Méthodes et instruments	8
3) Entretiens semi-structurés.....	9
4) Enquête en ligne	11
Revue de littérature	13
I. Observations générales sur les « communautés de recherche »	15
A. Définition des « communautés de recherche ».....	15
1) Appartenance ou non à une ou des communautés de recherche	15
2) Différents types de communautés	16
3) Un périmètre imprécis.....	18
B. Organisation de ces communautés	19
1) Mode d'organisation.....	20
2) Culture de partage	21
3) Lien entre la typologie de la recherche et l'organisation en communauté.....	22
C. Focus sur deux profils particuliers : les communautés de données et les chercheurs « sans données »	26
1) Les communautés de données	26
2) Les chercheurs « sans données ».....	27
II. Rôle de l'organisation en « communautés de recherche » pour les pratiques ORD	28
A. Les facteurs de motivation et les obstacles pour les chercheurs au développement de l'ORD	28
1) Les préoccupations ressenties par les chercheurs dans les pratiques ORD	28
2) Les obstacles au développement de l'ORD.....	30
3) Les facteurs de motivation au développement de l'ORD	33
B. L'utilisation de standards	35
C. Le niveau de compétences en ORD	40
III. Les leviers possibles pour le développement de l'ORD	46
A. Acteurs-clés.....	46
B. Stratégie d'implémentation	51
C. Sensibilisation, éducation et formation	53
IV. Recommandations pour développer le niveau de pratiques de l'ORD au sein des communautés	54
1) « Talking the same language and starting from the beginning »	54

2) Accompagner la construction et la consolidation des communautés de recherche.....	55
3) Développer les mesures d'incitation (amoindrir les obstacles).....	55
4) Impliquer les différents acteurs et formaliser leur contribution.....	55
5) Utiliser la communication et la formation en temps opportun	55
6) Fournir le bon niveau de soutien pour la définition et l'adoption des normes.....	56
7) Planifier un financement durable	56
Matrice de décision	57
Conclusion	59
Bibliographie.....	60
V. Annexes.....	61
Annexe 1: Concepts-clés	61
Annexe 2: Modèle de guide pour les entretiens semi-structurés.....	62
Annexe 3 : Thèmes et sous-thèmes de la grille d'analyse des données collectées avec les entretiens	66
Annexe 4: Contenu détaillé de l'enquête en ligne	67
Annexe 5 : Liste des sites web issus des réponses à l'enquête.....	96
Annexe 6 : Liste de sites ou pages Internet rattachées à des communautés de données	98

Table des figures

Figure 1: Chronologie du travail	9
Figure 2: Structure principale de l'enquête.....	11
Figure 3: Institution d'affiliation.....	12
Figure 4: Domaines de recherche.....	13
Figure 5: Appartenance à une communauté de recherche	16
Figure 6 : Caractérisation des communautés de recherche.....	17
Figure 7 : Domaine ou discipline de recherche	18
Figure 8 : Institution d' affiliation.....	19
Figure 9 : Organisation des communautés de recherche	20
Figure 10 : Envergure typique de la recherche	23
Figure 11 : Echelle de temps typique de la recherche	24
Figure 12 : Nature de la recherche.....	25
Figure 13 : Type de données	25
Figure 14 : Taille des données	26
Figure 15 : Inquiétudes concernant les pratiques ORD.....	29
Figure 16 : Obstacles pratiques concernant le partage et la réutilisation des données	32
Figure 17 : Facteurs de motivation pour les pratiques ORD	34
Figure 18 : L'utilisation de standards pour les données.....	36
Figure 19 : L'utilisation de standards pour les métadonnées	37
Figure 20 : L'utilisation de 'PIDs'	37
Figure 21 : L'utilisation de standards de format de fichiers.....	38
Figure 22 : L'utilisation de standards pour la documentation	39
Figure 23 : L'utilisation de standards pour les licences.....	39
Figure 24 : L'utilisation d'autres standards	40
Figure 25 : Élaboration d'un Data Management Plan (DMP)	42
Figure 26 : Préparer vos données pour les télécharger dans un dépôt	42
Figure 27 : Téléchargement des données de recherche dans un dépôt	43
Figure 28 : Partage des données de recherche via un dépôt.....	43
Figure 29 : Recherche de données de recherche produites par d'autres	44
Figure 30 : Réutilisation des données de recherche produites par d'autres	44
Figure 31 : Reproduction de résultats de recherche produits par d'autres.....	45
Figure 32 : Matrice de décision	58

Table des tableaux

Tableau 1 : Détails comparatifs sur les entretiens semi-structurés et l'enquête	9
Tableau 2 : Détails sur les entretiens semi-structurés	9
Tableau 3 : Liste des inquiétudes avec extraits des entretiens.....	28
Tableau 4 : Liste des obstacles avec extraits des entretiens.....	31
Tableau 5 : Liste des facteurs de motivation avec extraits des entretiens	33
Tableau 6 : Acteurs-clés pour l'ORD	46
Tableau 7 : Identifiants des recommandations pour la matrice de décisions	57

Sigles et acronymes

Acronyme	Désignation
DaSCH	Data and Service Center for the Humanities
DH+	Digital Humanities
DLCM	Data Life Cycle Management
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR	Findable, Accessible, Interoperable, Reusable
FHGR	Fachhochschule Graubünden
GDR	Gestion des données de la recherche
HEG	Haute école de gestion
OA	Open Access
ORD	Données de recherche ouvertes
OS	Open Science
PID	Persistent Identifier
RDM	Research Data Management
SIB	Swiss Institute of Bioinformatics
ZHdK	Zürcher Hochschule der Künste

Résumé

Ce rapport est issu de l'étude mandatée par swissuniversities auprès de la Haute école de gestion de Genève. Se focalisant sur les communautés de recherche, il caractérise les pratiques et les besoins des chercheurs¹ en Open Research Data (ORD) et contribue à orienter la stratégie de soutien à l'ORD. Il met en lumière les définitions variées de la « communauté de recherche » formulées par les chercheurs, ainsi que leur méconnaissance quant à leur appartenance à une communauté de données. Ce rapport démontre également qu'il n'y pas de modèle clair des communautés de recherche et que ces dernières disposent d'une organisation informelle. Il relate par ailleurs les impacts de l'appartenance à une communauté de recherche sur les pratiques ORD et met en avant les résultats suivants :

- les préoccupations, obstacles et motivations pour l'ORD ne sont pas significativement différents selon l'appartenance ou non à une communauté
- les standards sont plus largement utilisés par les scientifiques appartenant à une communauté
- les compétences sont plus étendues pour ces derniers ; le niveau de développement de ces compétences est tributaire en partie du soutien d'une communauté, d'une institution ou d'une infrastructure

Au vu de ces deux derniers résultats, l'utilisation de standards et le niveau de développement des compétences semblent être de bons candidats pour l'évaluation du niveau de maturité d'une communauté.

Le rapport présente par ailleurs des leviers possibles pour la création et le développement d'une communauté de recherche, à savoir :

- l'engagement d'acteurs clés
- l'adoption simultanée d'approches bottom up et top down
- la sensibilisation, l'éducation et la formation

Enfin, il formule les recommandations suivantes :

- définir le concept de « communauté de recherche »
- accompagner la construction et la consolidation des communautés de recherche
- fournir le niveau de soutien adéquat pour la définition et l'adoption de normes

Le niveau de priorité pour chaque recommandation est représenté de façon synthétique sous la forme d'une matrice de décision, en fonction du degré de sévérité de la recommandation (son niveau d'importance), le besoin du groupe cible et l'effort nécessaire pour son implémentation.

¹ Pour faciliter la lecture de l'ensemble du rapport, nous employons le masculin comme genre neutre afin de désigner aussi bien les hommes que les femmes.

Introduction et contexte

Dans le cadre de la Stratégie nationale pour des données de recherche ouvertes initiée par sa Délégation Open Science, swissuniversities a mandaté la HEG et la FHGR pour réaliser deux études complémentaires : (1) caractériser le paysage des données de recherche ouvertes (ORD) en Suisse (Erstellung einer Übersicht über die ORD Landschaft in der Schweiz) ; (2) caractériser les pratiques et les besoins des chercheurs en ORD en se focalisant sur les communautés de recherche (ORD-Fähigkeit von wissenschaftlichen Gemeinschaften). Cette deuxième étude a été réalisée par la HEG entre mars et septembre 2021. Ce document en est le rapport final.

Les résultats obtenus devraient contribuer à orienter la stratégie de soutien à l'Open Research Data (ORD). En caractérisant les pratiques et les besoins actuels des chercheurs, et en distinguant ceux-ci par communauté et éventuellement leur granularité par communauté, ils permettront d'optimiser les ressources consacrées à l'Open Research Data, ainsi que de consolider la stratégie et le plan d'action qui en découle.

Les données collectées pour réaliser ce travail l'ont été en deux phases. Des entretiens semi-structurés de représentants d'infrastructures de partage de données et de communautés de recherche ont d'abord été réalisés. Puis les chercheurs académiques suisses ont été consultés au moyen d'un questionnaire en ligne.

La suite de ce document présente la méthodologie de recherche adoptée, puis la revue de littérature autour des thèmes et des concepts-clés de ce travail. La définition de ces concepts est donnée dans l'Annexe 1. Il détaille ensuite les résultats obtenus avec la collecte des données et présente leur synthèse ainsi que des recommandations.

Méthodologie

1) Objectifs

Les principaux objectifs de ce travail sont les suivants :

1. Caractériser le niveau de maturité de la communauté de recherche à l'égard de l'ORD
2. Identifier les besoins des communautés en termes d'appui à l'amélioration de leurs pratiques et de leur culture ORD
3. Caractériser la granularité de ces besoins par communauté
4. Préciser les acteurs de l'ORD à impliquer, par catégories de besoins et éventuellement par communauté
5. Faire des recommandations pour soutenir les communautés dans leur pratique ORD

Ces objectifs visent à répondre aux questions de recherche suivantes :

1. Comment évaluer le niveau de maturité en ORD de la recherche académique en Suisse ? A partir de quels critères ? Quel est ce niveau de maturité ? Quelle différenciation existe-t-il entre les communautés ? Que montrent ces résultats sur la culture de partage des données dans les différentes communautés ? Que révèlent-ils sur l'étendue des besoins pour atteindre un niveau complet de pratique ORD pour les différentes communautés ?
2. Quels sont les critères pour évaluer les besoins spécifiques pour l'avancement des communautés en ORD ? Quels sont ces besoins ?
3. Les besoins sont-ils tous les mêmes au sein des communautés ou y a-t-il des différences au niveau des sous-communautés ? Jusqu'à quel sous-niveau (pour chaque communauté) ?
4. Qui sont les acteurs qui doivent être impliqués ? Est-il nécessaire de différencier les communautés en ce qui concerne ces parties prenantes ? Comment peuvent-ils contribuer, à l'échelle de l'institution, à accroître et à améliorer le partage des données de recherche ?
5. Quel type de soutien aux communautés sera le plus efficace pour améliorer leurs pratiques ORD ? Quelles autres recommandations sur ce soutien peuvent être faites ?

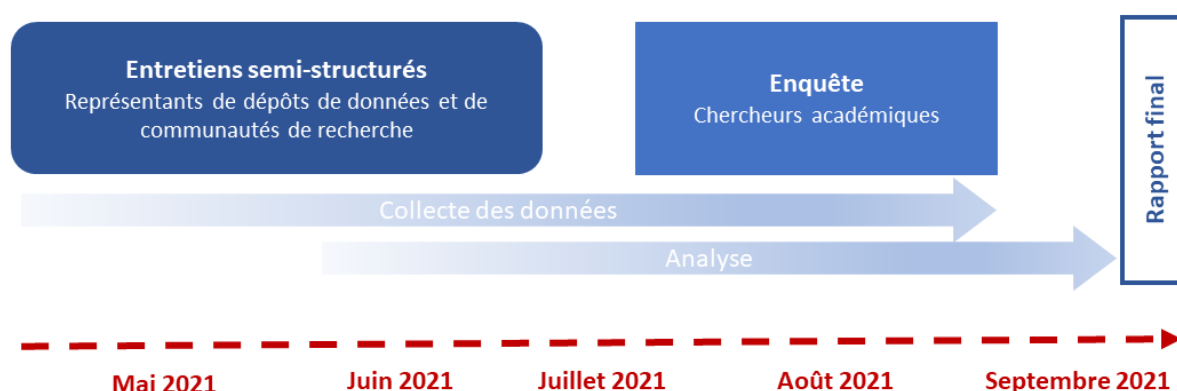
2) Méthodes et instruments

Nous avons utilisé successivement deux méthodes principales pour collecter les informations:

- des entretiens semi-structurés auprès des dépositaires de données et des représentants de la communauté
- une enquête en ligne adressée aux chercheurs universitaires suisses

La Figure 1 ci-dessous illustre la chronologie générale du travail, avec les phases successives et quelque peu superposées de collecte, d'analyse et de synthèse des données pour la production du rapport final.

Figure 1: Chronologie du travail



Les objectifs des deux méthodes étaient de collecter des informations sur les pratiques et la culture actuelles en matière de ORD, les besoins d'amélioration des pratiques en matière de ORD et les contributeurs à impliquer la communauté des chercheurs.

Avec les entretiens semi-structurés, nous avons cherché à obtenir des premiers éléments sur le sujet ainsi que des informations utiles pour aider à la conception de l'enquête (structure, contenu). Avec l'enquête en ligne, nous avons cherché à obtenir des résultats globaux et significatifs à l'échelle académique suisse et à l'échelle de la communauté.

Les détails comparatifs des deux méthodes sont donnés dans le Tableau 1.

Tableau 1 : Détails comparatifs sur les entretiens semi-structurés et l'enquête

Méthode	Public cible	Caractéristiques
Entretiens semi-structurés	Gestionnaires de dépôts de données et représentants des communautés	Flexible : guide d'entretien adapté à chaque personne interrogée à partir d'un modèle commun ; informations riches ; taux de réponse de 100%
Enquête	Chercheurs académiques suisses	Fournit suffisamment d'informations pour tirer des conclusions significatives ; peu d'espace pour enregistrer les informations non prévues ; taux de réponse inférieur à 100%

3) Entretiens semi-structurés

Les détails concernant les personnes interrogées et les dates des entretiens sont donnés dans le Tableau 2.

Tableau 2 : Détails sur les entretiens semi-structurés

Nom	Poste	Sujets	Date
Pierre-Yves Burgi	Initiateur du projet OLOS, codirecteur du DLCM	Publication et préservation de la RD, services RDM	21/05/21

Daniela Subotic and Rita Gautschy	Responsable du service clientèle/gestion des produits de DaSCH	Référentiel de données disciplinaires et services : DasCH (sciences humaines et sociales)	27/05/21
Torsten Schwede	Vice-président de la recherche, Université de Bâle, chef du groupe de recherche Computational Structural Biology Group, Institut suisse de bioinformatique)	Culture, pratiques et besoins ORD en matière de données bio-informatiques	28/05/21
Georg Lutz	Directeur de FORS	Référentiel de données disciplinaires et services : FORSbase (sciences sociales et politiques)	17/06/21
Beate Böckem and Susanne Schumacher	Responsable de la recherche à la Haute école des arts de Zurich et coprésidente du conseil numérique de la ZHdK, responsable du programme Digital council	La culture, les pratiques et les besoins des communautés de recherche du ZHdK	02/07/21
Claire Clivaz (contribution écrite envoyée par courriel)	Responsable des DH+ au SIB	Culture, pratiques et besoins du groupe de recherche DH+ et des communautés de recherche connexes.	22/08/21

Le guide d'entretien a été fourni à la ou aux personnes interrogées quelques jours avant la date de l'entretien. Ce guide d'entretien suivait la même structure pour tous (voir annexe 2) mais était adapté en fonction de la position spécifique de la personne interrogée. Tous les entretiens ont été réalisés en ligne et ont été enregistrés pour un usage interne uniquement, sous réserve de l'accord des personnes interrogées.

Le rapport d'entretien a ensuite été produit et envoyé quelques jours après l'entretien pour validation ou consolidation avec les commentaires de la personne interrogée.

En plus de ces cinq entretiens, Claire Clivaz, responsable des Humanités numériques au SIB, nous a également envoyé ses réponses écrites à des questions similaires à celles posées lors des entretiens semi-structurés.

Certaines informations recueillies lors des entretiens ont été utilisées pour concevoir la structure et formuler les questions de l'enquête de manière adéquate.

L'analyse des données collectées a été réalisée au moyen d'un tableau spécifique où le contenu pertinent de chaque rapport final d'entretien a été collé et organisé par thème (voir la liste de ces thèmes dans l'annexe 3).

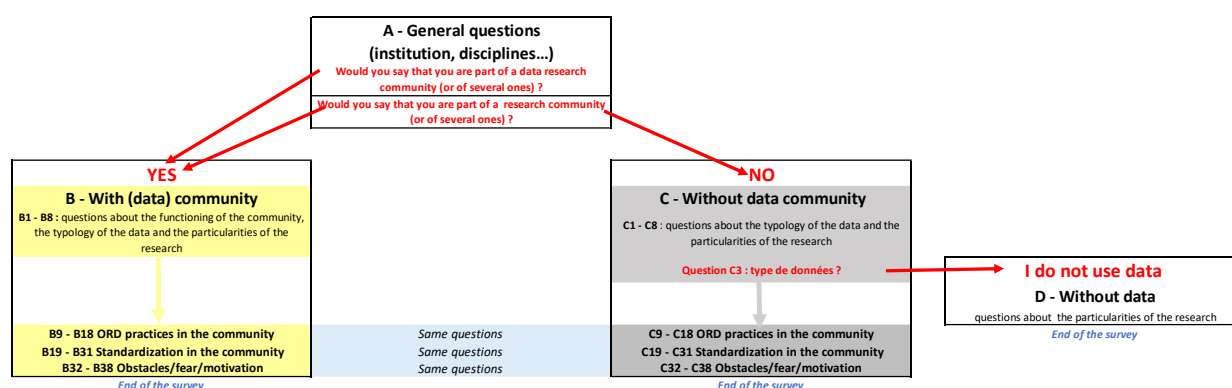
4) Enquête en ligne

Contenu et structure

Le contenu de l'enquête est un mélange de questions fermées avec des listes déroulantes de réponses possibles et quelques questions ouvertes (pour des informations supplémentaires). L'enquête a été conçue avec Qualtrics, qui permet une analyse approfondie des réponses recueillies.

La structure de l'enquête est illustrée sur la Figure 2.

Figure 2: Structure principale de l'enquête



La section A contient des questions sur l'affiliation à la recherche et le domaine de recherche de la personne interrogée. A la fin de cette section, il est également demandé au chercheur s'il appartient à une communauté de recherche, et si oui, à quel type de communauté de recherche (une communauté liée aux données, à un projet, à un laboratoire de recherche, à une discipline ou un domaine de recherche, à une infrastructure).

Si le chercheur appartient à une communauté, il est dirigé vers la section B. Sinon, il est dirigé vers la section C. Les premières questions de ces deux sections portent alors sur les caractéristiques de la communauté de recherche et le type de recherche qu'elle effectue. Des questions supplémentaires au début de la section B permettent également de caractériser la communauté.

La deuxième partie de chaque section (questions de B9 et de C9) est similaire et traite des points suivants:

- les pratiques ORD dans la communauté
- l'attitude de la communauté par rapport aux normes : l'existence éventuelle et l'utilisation des normes
- les obstacles, les préoccupations et les motivations concernant les pratiques de ORD

Le contenu complet de l'enquête est donné dans l'annexe 4.

Calendrier

L'enquête a été ouverte le 21 juillet, avec un appel formel à contributions envoyé à la liste swiss-lib, au réseau informel RD CH via son coordinateur, et à une liste d'experts OS/ORD et d'acteurs principaux de l'ORD.

Un deuxième appel à contribution a été envoyé le 21 août, avec un appel spécifique adressé aux chercheurs des Hautes Ecoles Spécialisées (qui affichaient à l'époque un taux de réponse bien inférieur aux autres affiliations). L'enquête a été clôturée le 8 septembre.

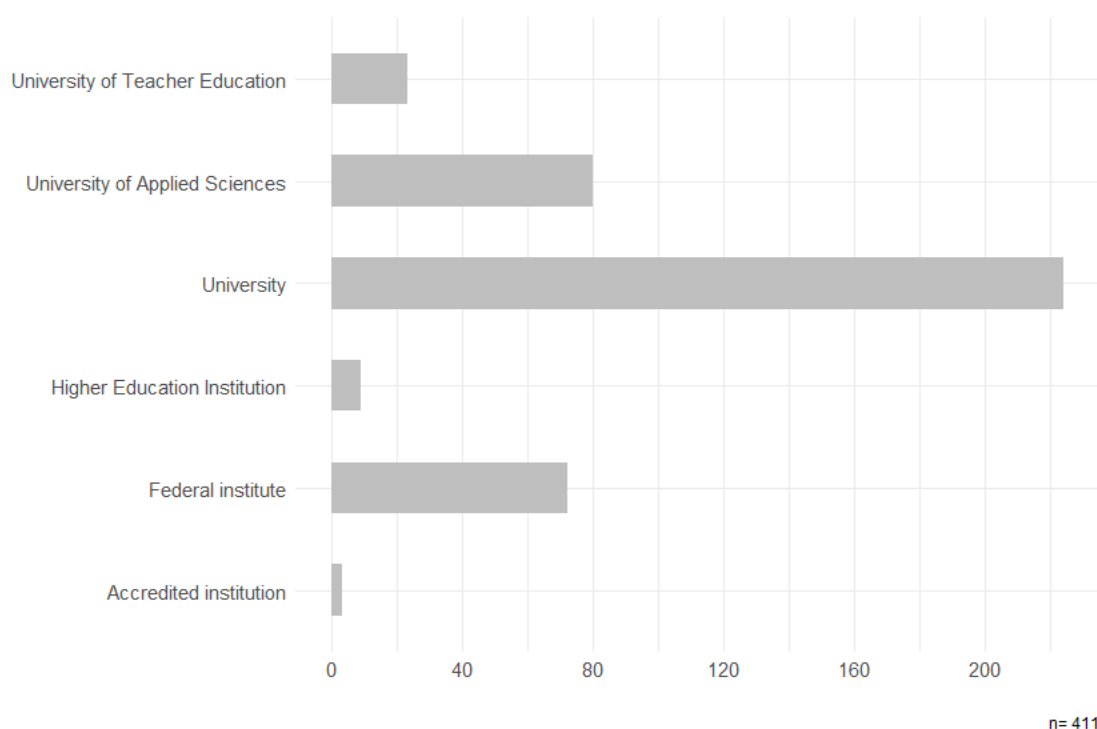
Contributions

Au total, 185 réponses complètes ont été recueillies, ce que nous considérons comme un nombre suffisamment important pour tirer des résultats généraux, du moins pour les questions communes à tous (section A principalement). Certains chercheurs n'ont répondu qu'à un sous-ensemble des questions, ce qui signifie que le nombre de réponses varie en fait en fonction de la question. Par exemple, ce nombre s'élève à plus de 416 pour la première question sur l'institution d'affiliation, à 356 pour la question sur le domaine ou la discipline de recherche et à 200 pour la question relative à la description de la communauté de recherche. Nous avons décidé d'inclure de toute façon ces réponses dans l'analyse et les résultats, à condition que le nombre total de réponses soit pris en compte dans l'analyse et soit visible sur les graphiques de résultats s'ils sont inclus dans le rapport.

Environ deux tiers des contributeurs ont déclaré appartenir à une communauté de recherche et ont donc été dirigés vers la section B de l'enquête après avoir rempli la section A. Pour des raisons de significativité des résultats, nous avons décidé de rediriger vers la section C les contributeurs déclarant ne pas appartenir à une communauté de recherche et ceux déclarant ne pas savoir (s'ils appartiennent à une communauté de recherche). Dans un souci de simplicité, nous désignons ce groupe de contributeurs comme des personnes ne faisant pas partie d'une communauté, ou des non-membres d'une communauté.

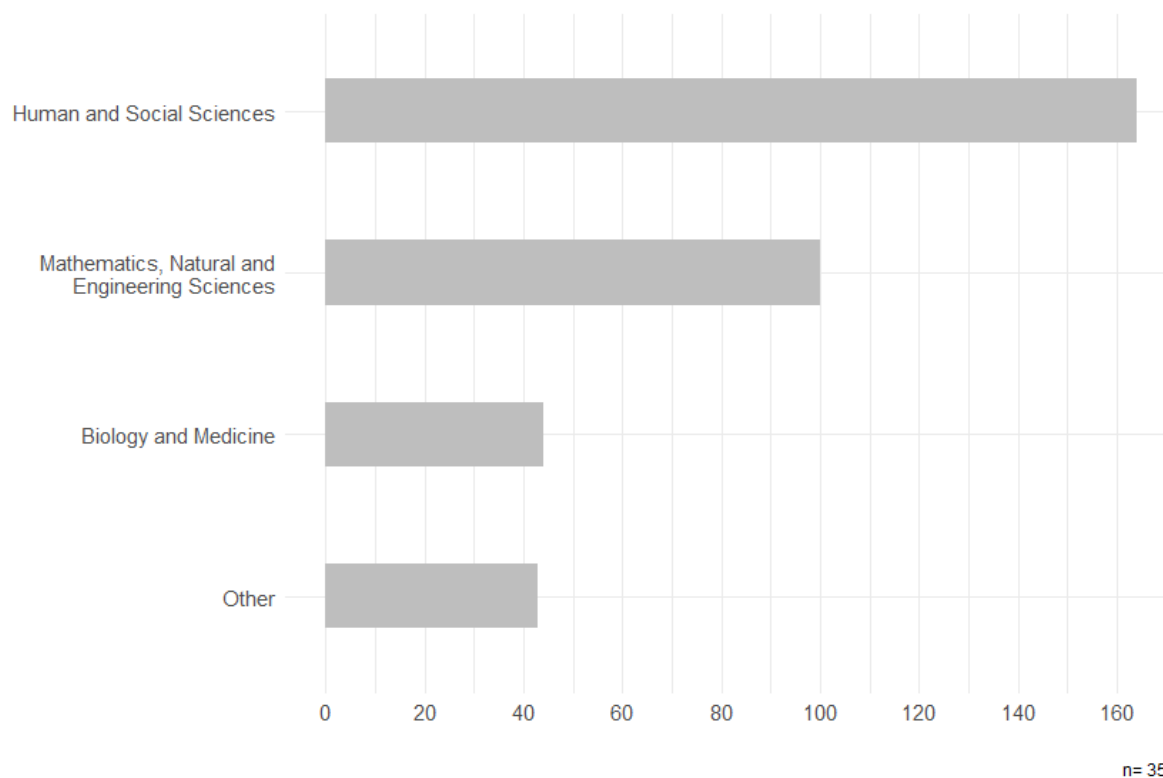
Bien que cette enquête ne s'adresse qu'aux chercheurs académiques suisses (ceci a été précisé dans la partie introduction et dans le texte de l'appel à contribution), deux bibliothécaires ont également contribué. Leurs réponses ont pu être facilement identifiées puisqu'ils ont indiqué leur position dans la réponse à une question.

Figure 3: Institution d'affiliation



La plupart des contributions (53% des réponses) proviennent de chercheurs universitaires (Figure 3). Parmi ceux-ci, l'Université de Fribourg est surreprésentée avec 76 contributions (18% du nombre total de contributions).

Figure 4: Domaines de recherche



Field of interest	n
Human and Social Sciences	164
Mathematics, Natural-and Engineering Sciences	100
Biology and Medicine	44
Other	43

La répartition des chercheurs par champ disciplinaire (Figure 4) est plus homogène que celle de l'affiliation institutionnelle, bien que leur nombre soit plus important en Sciences humaines et sociales (44% des réponses).

Revue de littérature

Dans sa *Stratégie Nationale Suisse Open Research Data*, swissuniversities souligne la nécessité de « mettre en place et de développer le paysage ORD conformément aux besoins des chercheur-e-s et dans le respect des conditions cadres politiques » (swissuniversities 2021, p. 20).

Berman et al. (2013) pointent quant à eux l'importance de la collaboration d'une communauté dans l'émergence et la réussite de pratiques ORD. En effet, « *The role of community in shaping the [Protein Data Bank] and helping it to evolve into a vital resource for biological research cannot be underestimated* » (Berman et al. 2013, p.1).

Dans une veine similaire, Copper et Springer expliquent que « *the best policies and standards achieve little without buy-in from the research community* » (Cooper, Springer 2019, p. 5)

Selon Berman et al. (2013), la définition d'une communauté de données peut évoluer avec le temps. Dans le cas de la Protein Data Bank, le profil des utilisateurs s'est en effet diversifié et « *the user community has expanded from the small group of activists who saw the need for an archive to their research to an ever-larger group or users who far outnumber the depositors* » (Berman et al. 2013, p.6).

Pour Cooper et Springer (2019, p. 4), les communautés de données sont des « *formal or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries* ».

Concernant la mise en place d'une pratique ORD, Swissuniversities établit dans sa *Stratégie Nationale Suisse Open Research Data* (2021, pp. 8-10) quatre objectifs :

1. Soutien aux chercheur-e-s et aux communautés de recherche pour concevoir des pratiques ORD et pour les adopter.
2. Développement, promotion et maintenance d'infrastructures et de services de base financièrement viables pour l'ensemble des chercheur-e-s :
 - Infrastructures et services de base
 - « Data Stewards »
3. Développement des compétences des chercheur-e-s et échange de bonnes pratiques
4. Développement des conditions systémiques et favorables pour les institutions et les communautés de recherche :
 - Mécanismes de récompenses à modifier
 - Accès aux supports et ressources pour comprendre les défis juridiques, éthiques et sociaux relatifs à la production, l'accès et l'utilisation des données
 - Politiques et mesures d'implémentation doivent s'aligner aux standards et actions internationales

Selon Cooper et Springer (2019), le partage des données et ainsi les pratiques ORD sont indissociables des communautés de données (data communities). Les auteurs mettent en avant trois caractéristiques pour leur succès : un développement ascendant (bottom-up development), l'absence ou l'atténuation d'obstacles techniques et des standards.

En ce qui concerne les standards, le document *FAIR Data Maturity Model* (Research Data Alliance FAIR Data Maturity Model Working Group 2020) met en avant des préconisations pour une utilisation FAIR des données, à savoir des données découvrables, accessibles, interopérables et réutilisables, par les machines comme les humains. Il est ainsi possible de faire émerger notamment les standards suivants:

- Standards de métadonnées
- PIDs
- Standards de licence
- Standards de format de fichier
- Standards sur la documentation

Cooper et Springer (2019) évoquent également la question des standards et des principes FAIR pour un partage de données réussi: « *[Many] data sharing advocates have embraced the FAIR data principles [...] as the standard benchmark for data sharing success* » (Cooper, Springer, 2019, p.2).

I. Observations générales sur les « communautés de recherche »

Comme observé dans la revue de la littérature, les communautés de recherche sont identifiées comme des acteurs clé pour le développement de l'OS et de l'ORD en particulier. La stratégie de swissuniversities s'aligne logiquement sur ce constat. Lors des entretiens, l'importance donnée à ces communautés de recherche et la nécessité de les impliquer dans le développement de l'ORD semblent également unanimes.

« You need to invest a lot in community buildings. You have to understand all these projects as community-building projects and that is the hard part. » (Lutz 2021)

« One of the precondition for OS to work is that we have understanding for the motivations and the limitations of all players in the communities. Otherwise, there is no motivation to share anything. » (Ackermann, Schwede 2021)

Cependant, malgré ce constat partagé, le périmètre de ces communautés reste à définir et particulièrement dans l'environnement suisse. En effet, les communautés de recherche semblent correspondre à une idée plus transversale de la recherche avec des approches interdisciplinaires ou internationales.

« In the interest of establishing and passing on best practices, it is crucial to strengthen knowledge exchange among research communities through the creation of networks between institutions and between data stewards or similar positions within higher education and research institutions. » (Swissuniversities 2021)

Si des définitions existent dans la littérature, il est important de s'accorder également dans la pratique, surtout dans la perspective de s'appuyer sur ces communautés pour développer la culture et les pratiques en ORD.

A. Définition des « communautés de recherche »

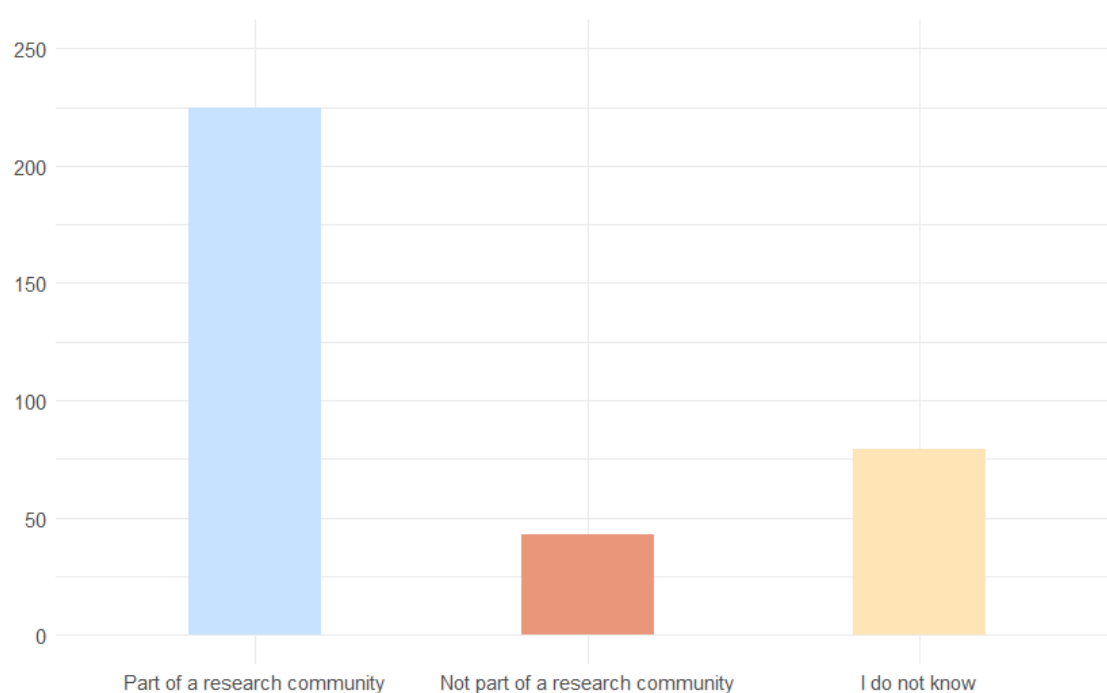
La sémantique autour de la notion de « communauté de recherche »² ne semble pas encore stabilisée. Un des enjeux de ce mandat était de s'assurer de la concordance du discours entre la théorie et la pratique.

1) Appartenance ou non à une ou des communautés de recherche

La notion de « communauté de recherche » est présente sur le terrain. Pendant les entretiens, cette notion a été largement utilisée dans les échanges sans que l'existence tangible de ces communautés dans l'environnement de la recherche en Suisse ne soit remise en question. L'enquête a confirmé cette réalité avec les réponses à cette question « *Would you say that you are part of a research community (or of several ones)?* »

² Dans ce travail, le terme « communauté de recherche » a été choisi pour traduire l'idée de « research community »

Figure 5: Appartenance ou non à une communauté de recherche



n= 347

Comme le montre la Figure 5, 230 chercheurs soit 65% des chercheurs interrogés déclarent appartenir à une communauté de recherche. La précision sur la possibilité d'appartenir à plusieurs communautés a été initialement ajoutée afin de limiter les réponses négatives basées sur le singulier de la question. Or, il s'est avéré qu'un très grand nombre de chercheurs citent plusieurs communautés quand il leur est demandé de préciser à quelle(s) communauté(s) ils appartiennent. On pourrait facilement conclure que le fait d'appartenir à une communauté de recherche augmente la possibilité d'appartenir à plusieurs communautés. Cependant, la vision de ces communautés de recherche semble assez floue comme le montre les 23% de chercheurs qui répondent ne pas savoir s'ils sont dans une communauté ou pas. Cette vision trouble entraîne des confusions qui limitent l'interprétation d'une partie des réponses. Cette problématique sémantique avait été repérée lors des entretiens et donc investiguée dans l'enquête.

2) Différents types de communautés

Les entretiens ont offert l'opportunité de proposer une première vision de ces communautés de recherche avec un axe marqué sur les données, qui faisait logiquement sens dans une réflexion sur les pratiques ORD. Dans le guide d'entretien (Annexe 2), la question « *Among your institution users [or infrastructure users], do you identify specific research data communities ?* » était accompagnée de la définition suivante :

« By research communities, we mean networks of researchers sharing the same (type of) data, the same data sharing tools and infrastructures, the same ORD practices, and the same culture of Open Science, but not necessarily being from the same research discipline ».

L'objectif était de voir si cette définition correspondait à ce qui était repéré sur le terrain et dans le cas contraire d'ouvrir la discussion. A l'issue des cinq entretiens, aucune des personnes interrogées ne contredit réellement la définition mais, à chaque fois des compléments semblent nécessaires car comme le précise Torsten Schwede, les communautés de recherche ont un « large spectre »

(Ackermann, Schwede 2021). Georg Lutz rappelle par exemple que beaucoup de chercheurs ne travaillent pas sur des données en tant qu'objets numériques :

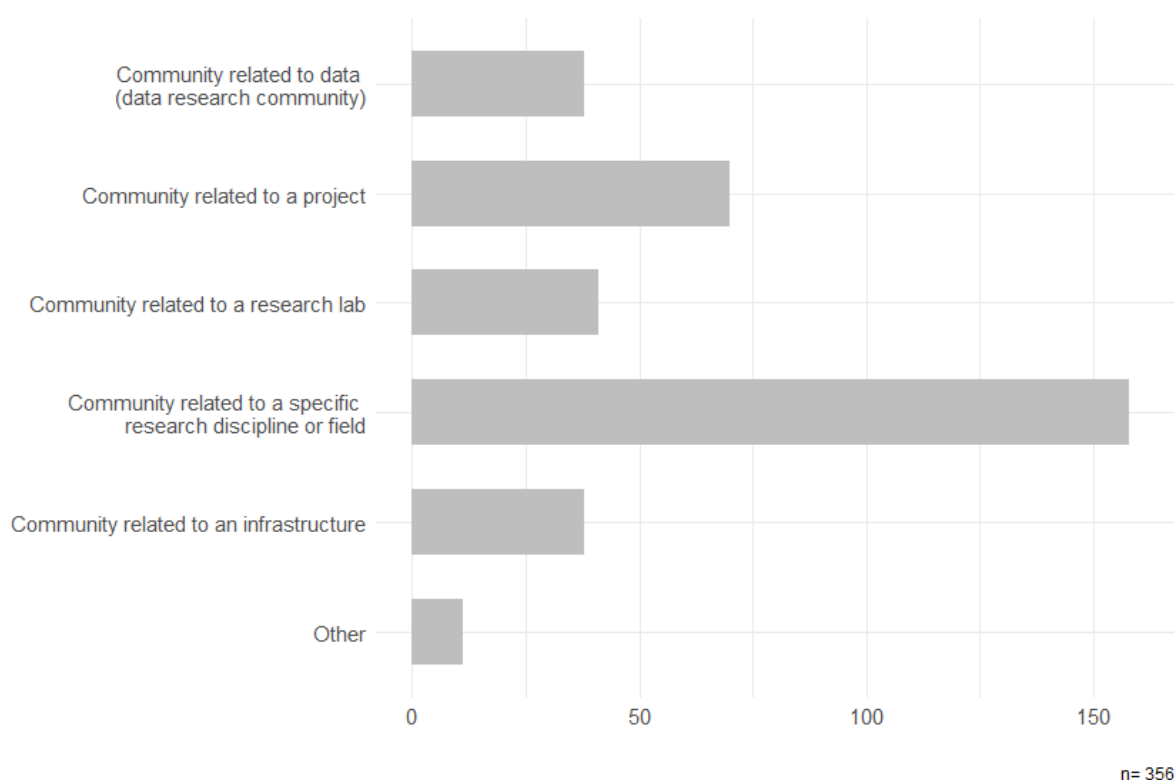
« Many researchers do not create data [...]. They use manuscripts, other kind of information than datasets that you can easily store ». (Lutz 2021)

Cette réalité, même marginale, limite, par exemple, la définition initiale centrée majoritairement sur l'utilisation de données. Globalement, durant les entretiens, les communautés sont décrites comme des groupes de chercheurs qui s'organisent autour de projets, d'infrastructures ou de méthodologies de recherche dans lesquels les données peuvent être similaires mais sans que cela semble être le point central.

« They are related to different kinds of infrastructures and to the very different and various kinds of research methodologies... It is very much not only discipline-related but also project-related. » (Böckem, Schumacher 2021)

On retrouve cette variété de création des communautés de recherche dans les résultats de l'enquête. En effet, 65% des chercheurs déclarent appartenir à une communauté de recherche mais la description qu'ils en font est variable.

Figure 6 : Caractérisation des communautés de recherche



Comme le reflète la Figure 6, les chercheurs qui ont répondu à l'enquête identifient leur communauté de recherche majoritairement en fonction d'une discipline ou d'un domaine de recherche. Le spectre des communautés de recherche reste pour autant visiblement large. En plus des définitions proposées dans la question, on trouve dans les réponses proposées par les chercheurs des communautés liées à la méthodologie de recherche, aux outils (logiciels) ou encore à la langue (réponse à la question de l'enquête sous forme de texte libre). Selon ces différents résultats issus de l'enquête, on constate que la définition de ces communautés reste variable.

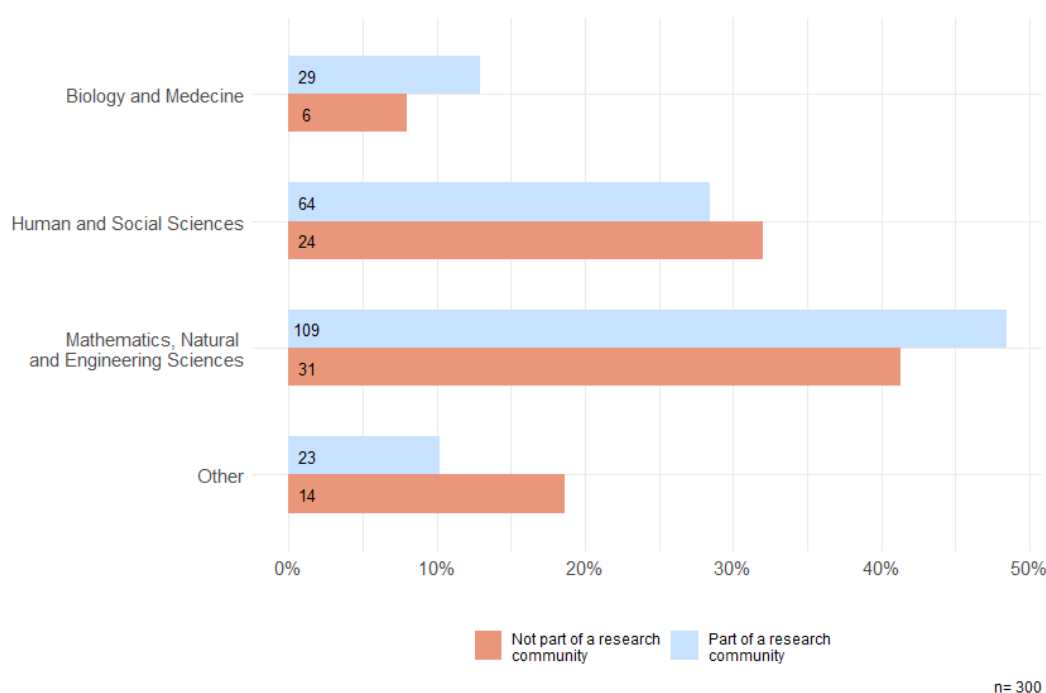
3) Un périmètre imprécis

Une partie significative des chercheurs interrogés (23%, 80 chercheurs, voir Figure 5) déclarent d'ailleurs ne pas savoir s'ils appartiennent à une communauté de recherche. De plus, parmi les 13% (44 chercheurs) qui répondent ne pas appartenir à une communauté de recherche, plusieurs notent explicitement qu'ils ne savent pas ce qu'est une communauté de recherche et dans ces mêmes 13% hors communautés de recherche, une quinzaine déclare paradoxalement demander l'aide de « leur communauté » pour certaines pratiques ORD. Tous ces éléments montrent que le périmètre des communautés de recherche est variable et même assez imprécis.

Si, comme dans la définition initiale, il semble important de préciser que les communautés ne sont pas uniquement disciplinaires, concrètement, l'approche disciplinaire reste majoritaire. D'ailleurs, lorsque les personnes interrogées pendant les entretiens sont invitées à donner des exemples de communautés de recherche et à les ordonner selon leur niveau de pratiques en ORD, ce sont essentiellement des disciplines qui sont citées.

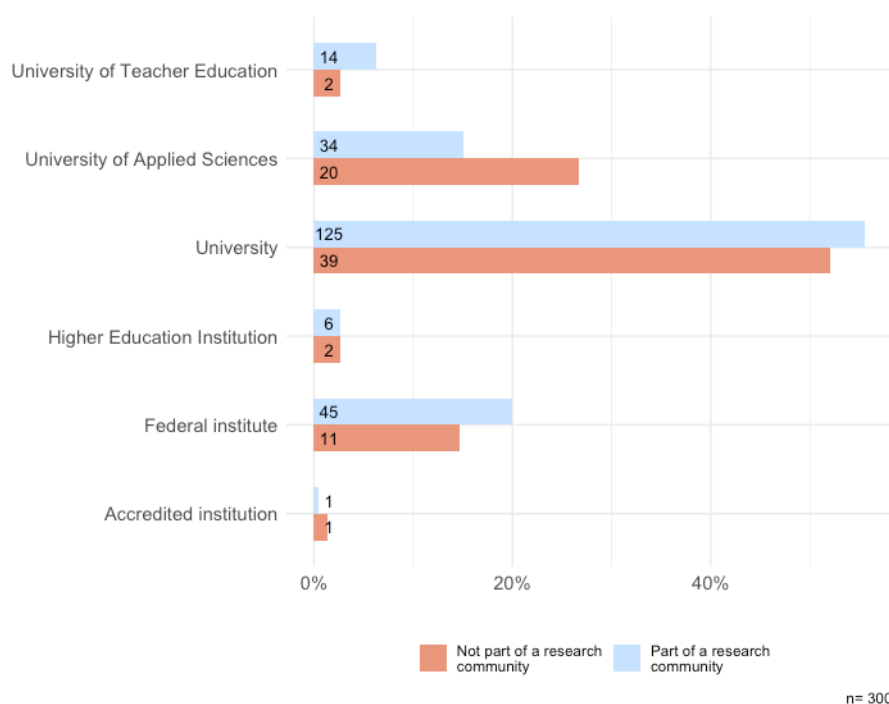
Pourtant, les résultats de l'enquête (Figure 7) montrent que le rattachement à une discipline particulière n'est pas lié de façon probante à la participation ou non à une communauté de recherche.

Figure 7 : Domaine ou discipline de recherche



Pour le dire autrement, la discipline ou le domaine d'études ne semble pas réellement influencer l'appartenance ou non à une communauté. De même le rattachement à une institution ne semble pas un élément concluant comme le montre la Figure 8.

Figure 8 : Institution d'affiliation



On peut éventuellement noter que les chercheurs des HES se disent (proportionnellement) moins rattachés à une communauté que ceux en universités. Ce léger constat peut être contrebalancé par ce commentaire issu de l'enquête :

« Le fait d'être associé à une Haute école spécialisée, avec les liens forts et les attentes entre la recherche et un domaine de pratique professionnelle, renforce l'appartenance à une communauté de recherche spécifique, dans ce domaine de pratique » (réponse à la question de l'enquête sous forme de texte libre)

La tentative de définition des communautés de recherche selon un périmètre institutionnel et/ou disciplinaire n'est donc pas probante. Pourtant, dans l'optique de travailler avec ces communautés et même d'éventuellement les impulser, il est important de comprendre comment elles s'organisent.

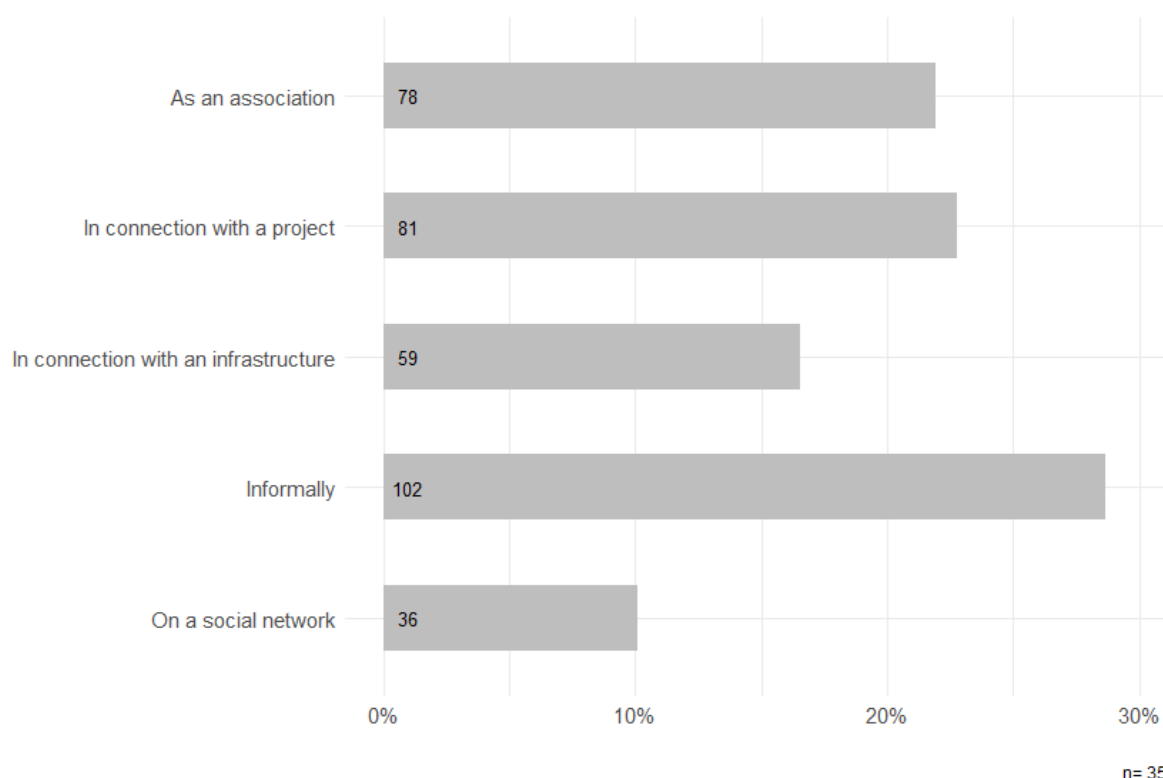
B. Organisation de ces communautés

Comme discuté ci-dessus, les communautés de recherche apparaissent dès les entretiens assez polymorphes. Pour cette raison, il nous semblait intéressant d'étudier si elles ont une existence factuelle et tangible, et d'éliminer ainsi l'hypothèse de simples réseaux informels de chercheurs qu'il serait beaucoup plus difficile d'atteindre et de soutenir dans leurs pratiques ORD.

Deux questions ciblent cet aspect dans l'enquête, une sur le mode d'organisation de la communauté et une autre sur l'existence d'un site web ou de pages propres à la communauté.

1) Mode d'organisation

Figure 9 : Organisation des communautés de recherche



La Figure 9 montre que le mode d'organisation des communautés est assez hétérogène. On retrouve logiquement une organisation autour de projets ou d'infrastructures, ceux-ci pouvant à la fois définir et structurer les communautés. Avec une proportion de réponses de 22%, l'association est un mode d'organisation significatif. Les réseaux sociaux sont aussi cités mais ne semblent pas centraux. Par contre, plus d'un quart de ces communautés sont organisées de façon informelle, ce qui les rend évidemment plus difficiles à atteindre et à mobiliser. Un chercheur commente ainsi cette question :

« In law, there is actually no talk of a "scientific community". Nevertheless, there are connections between colleagues in the same field in Switzerland and beyond the national context. These connections are partly informal, partly institutionalized. » (réponse à la question de l'enquête sous forme de texte libre)

On retrouve l'idée d'une appartenance à plusieurs communautés qui s'apparente plus à des réseaux informels sans délimitation réelle.

La question sur l'existence d'un site ou de pages web avait deux objectifs. Premièrement, il s'agissait de savoir si les communautés s'appuient sur au moins un outil tangible (ce qui peut déjà être un indicateur du développement de leur organisation interne) et deuxièmement, cela permettait de collecter des points d'accès à des informations complémentaires sur ces communautés (Annexe 5).

Si 28% répondent ne pas avoir de site web pour leurs communautés, la liste des pages web mentionnées par les chercheurs illustre la grande variété des modes d'organisation. En effet, de nombreuses pages sont hébergées sur des sites institutionnels et pointent directement vers des départements (ex : <https://www.unifr.ch/ecopol/fr/>), ou parfois vers des centres de recherche (ex : <https://www.zne.uzh.ch/en.html>). Certaines communautés, notamment celles organisées en

association, disposent de leur propre site (ex : <https://www.manep.ch/>). D'autres utilisent les réseaux sociaux (ex : <https://www.linkedin.com/groups/13763674/>).

On constate donc qu'il est difficile de trouver des structures « type » pour ces communautés. Même le nombre de membres se déploie de 5 à « *enormous to the point we can't count.* » (réponse à la question de l'enquête sous forme de texte libre).

Pourtant, si les communautés sont difficiles à délimiter, certains échanges informels semblent modeler certains comportements avec des accords tacites.

2) Culture de partage

En effet, certaines communautés semblent avoir des accords informels :

« Moreover, no structural biologist will review a paper which describes a structure not in the archive. No editor will even accept the paper. » (Ackermann, Schwede 2021)

Torsten Schwede les définit comme des « gentlemen agreements » et pour lui :

« It is also good for the disciplines to have a gentleman agreement (or other) in the community saying that nobody publishes before the first study author (the one who generated data) has published [...]. Otherwise, nobody would share. And that works surprisingly well. » (Ibid.)

Certaines communautés seraient plus enclines à partager leurs données que d'autres avec une forme de « culture du partage » plus développées.

« Certaines communautés (comme les sciences de l'environnement) sont beaucoup plus en avance sur le partage des données. Cela est aussi lié à la difficulté d'acquérir les données (efforts, temps, budget nécessaires). En physique des particules, par exemple, les machines comme les accélérateurs sont tellement chers que la tradition est de partager pour ainsi ne pas privilégier uniquement certains chercheurs. » (Burgi 2021)

L'exemple donné par Pierre-Yves Burgi révèle que cette habitude de partage des données est aussi liée à un contexte et des besoins précis. D'autres exemples issus des entretiens montrent que cela peut s'inscrire dans le temps :

« [Our group] does all steps around the Data life cycle [quite general in bioinformatics, and has been going on like that for a long time]: reuses, generates, processes, preserves and gives access »

« This is quite general in the domain of bioinformatics. This use and practice of the whole data lifecycle has been going on for a long time » (Ackermann, Schwede 2021)

Ou être également présent dans des communautés plus récentes :

« The younger communities are among the newly established communities. They rely stronger on infrastructures, tools or digital practices than communities like musical historians. » (Böckem, Schumacher 2021)

A contrario, un chercheur indique dans l'enquête ne pas faire partie d'une communauté de recherche car son domaine de recherche est nouveau (« *partly new research areas* »). La récence de la communauté ou de la discipline n'implique donc pas forcément cette culture du partage. Ce sont

plutôt les outils, les infrastructures utilisées qui semblent à l'origine de cette volonté de partager. A cela, on peut rajouter également la méthodologie de recherche qui se base sur une réutilisation des données, à l'instar de Torsten Schwede qui explique que « all his career has been based on reusing open data. ». Un autre exemple est donné par Beate Böckem et Susanne Schumacher :

« Other forms for reusing data, related to the very strong developed collaborative mode that is practiced in the arts: you are using an artistic method, or design method, and you take data, continue to write on them, to drawn on them, to program on them, to bring a new view on what is there and to document this, to bring this to a piece of art or a piece of programming » (Böckem, Schumacher 2021)

Concrétisation de cette culture du partage, les outils peuvent alors être pensés de manière interopérable et ouverts.

« In the field of computer music/electronic music: researchers at ZHdK are quite famous to contribute to community, to deploy patches or plug-ins for programs. They are always open source code that will be brought into community and reused in the community. » (Böckem, Schumacher 2021)

On constate que la formation d'une communauté et la culture du partage qui en découle semblent liées à des besoins initiaux dans l'utilisation de certains outils ou le choix de méthodologie. D'autres éléments factuels liés à la typologie de la recherche (envergure, échelle de temps typique et types de données) peuvent apparaître comme corrélés à une appartenance à une ou des communautés.

3) Lien entre la typologie de la recherche et l'organisation en communauté

Plusieurs thématiques liées à la typologie de la recherche sont apparues de manière récurrente dans les exemples donnés durant les entretiens : l'envergure, la durée de la recherche et la typologie des données. Elles semblent liées au fait d'appartenir ou non à une communauté. Par conséquent, ces thématiques ont été utilisées dans l'enquête pour vérifier si ces éléments sont liés à l'organisation en communauté.

Envergure de la recherche

La thématique de l'envergure de la recherche est liée à la nature des projets de recherche qui entraînent de plus en plus de partenariats au-delà des frontières. Les chercheurs impliqués dans ces projets européens ou internationaux doivent se conformer aux principes ORD de leurs partenaires comme le confirme, par exemple, Beate Böckem :

« In the national and international base, we are strongly related to the DORA declaration³, to the Florence principles⁴, to the Frascati manual⁵, so to all the international/national standards for quality and research » (Böckem, Schumacher 2021)

De ces partenariats découlent logiquement certaines pratiques :

« Practices that are related, fixed and rooted down to the national/international background, like the research catalogue⁶. » (Ibid. 2021)

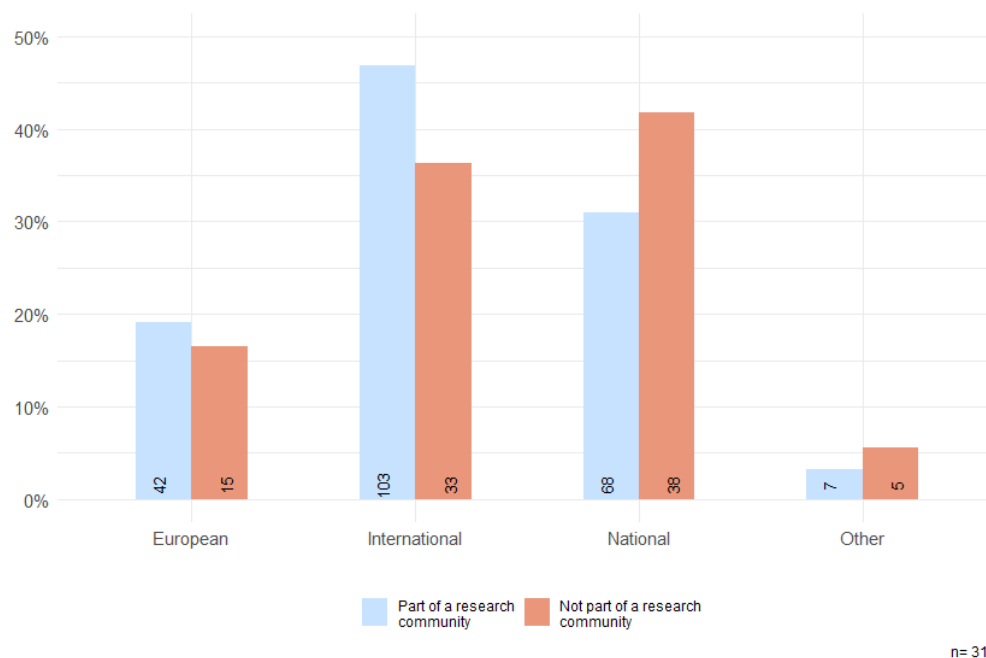
³ Déclaration de San Francisco sur l'évaluation de la recherche : <https://sfedora.org/read/read-the-declaration-french/>

⁴ The Florence Principles : <https://elia-artschools.org/page/Florence-Principles-On-the-Doctorate-of-the-Arts>

⁵ Manuel de Frascati : <https://www.oecd.org/fr/sti/manuel-de-frascati-2015-9789264257252-fr.htm>

L'envergure de la recherche semble influencer les pratiques des chercheurs mais plus largement l'appartenance à une communauté. Ainsi, Sonia Ackermann souligne l'envergure internationale des communautés: « *it is an international game because research communities are international.* » (Ackermann, Schwede 2021). En effet, les résultats de l'enquête confirment ce constat comme le montre la Figure 10 :

Figure 10 : Envergure typique de la recherche



n= 311

Les chercheurs qui déclarent faire des recherches au niveau international se disent plus souvent membres de communautés. Ce résultat s'étend aussi aux répondants faisant de la recherche à l'échelle européenne.

L'échelle de temps de la recherche

La durée et la planification dans le temps de certains projets de recherche sont mentionnées durant les entretiens et apparaissent plutôt comme un facteur qui rend difficile certaines collaborations :

« *Reward issue : who gets the paper? When can you publish ? Who gets credit ? Hard to solve if communities [collaborating on the same research project] work on different time scales.* » (Ackermann, Schwede 2021)

L'échelle de temps est également mentionnée par Rita Gautschy pour les recherches en archéologie comme un facteur influant sur l'envie de partager :

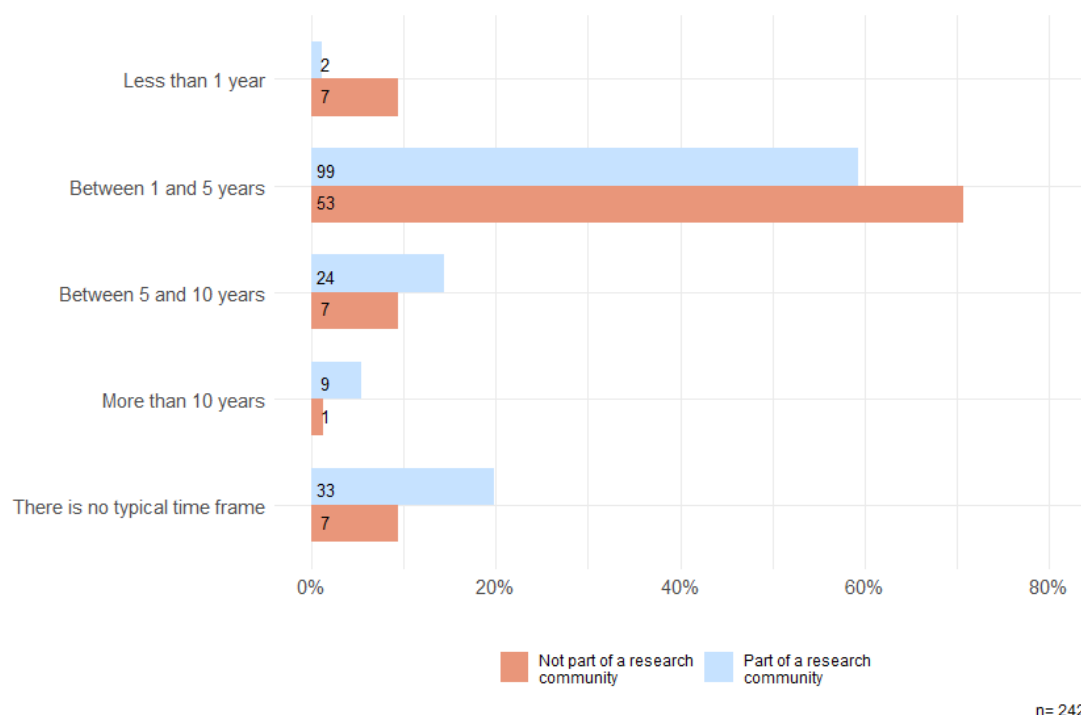
« *With Archaeology, there are very long time projects. The projects are often never-ending. When a project is really finished after a few years, people are more willing to share their data because it is finished for them too.* » (Gautschy, Subotic 2021)

Selon Georg Lutz, ce pourrait aussi être un prétexte pour ne pas partager ses données (quel que soit le domaine de recherche).

⁶ Research catalogue (an international database for artistic research) : <https://www.researchcatalogue.net/>

Cet exemple concernant l'archéologie illustre le fait qu'un temps de recherche très long peut limiter l'envie de partager. Les résultats de l'enquête, quant à eux, tendent à montrer que les chercheurs qui déclarent avoir des projets de recherche plus longs que cinq ans ont tendance à être plus fréquemment membre de communautés (Figure 11).

Figure 11 : Echelle de temps typique de la recherche



Ce résultat s'inverse si l'échelle de temps typique est inférieure à 5 ans.

Cependant, dans les commentaires de l'enquête, on observe cette remarque d'un chercheur en Pollination ecology :

« Data may not be public until all the research groups involved have published, which may take years ». (réponse à la question de l'enquête sous forme de texte libre).

Comme pour l'archéologie, la durée du projet rend difficile d'envisager une réelle culture du partage.

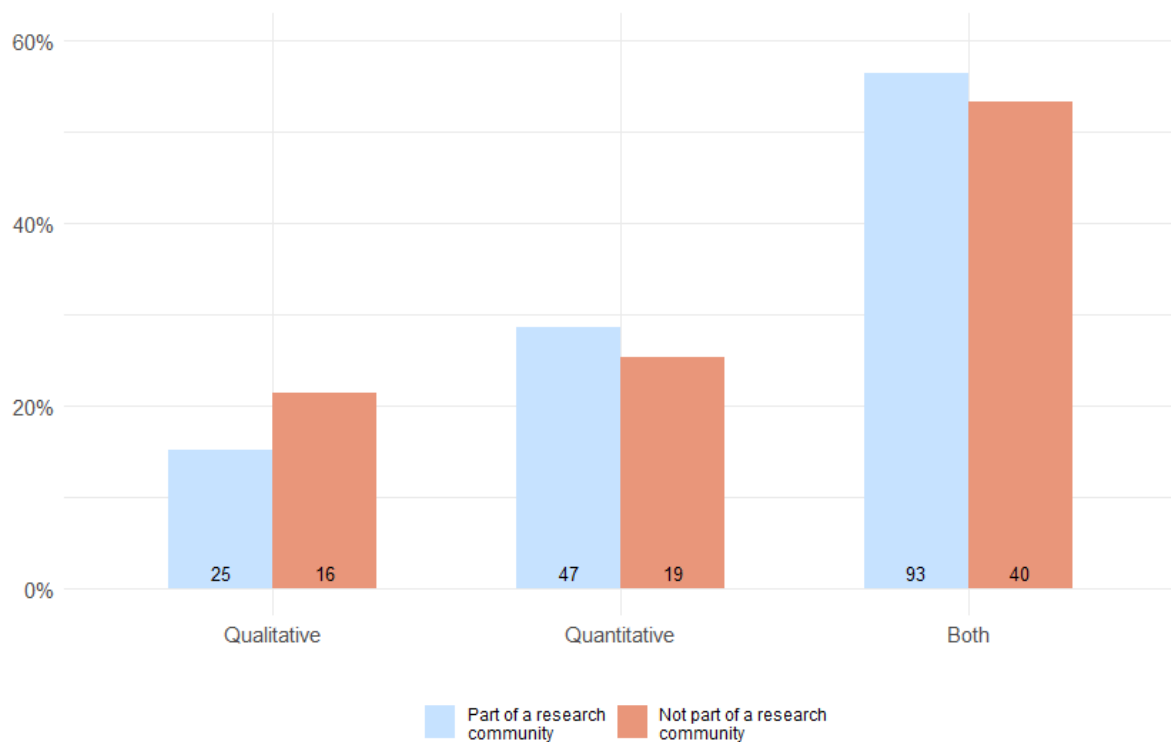
Type de recherche et typologie des données

Le type de recherche (qualitative, quantitative) et la typologie des données utilisées sont d'autres éléments abordés lors des entretiens et dans l'enquête.

« In qualitative research, people may not have yet a culture of data sharing (and data are super hard to anonymize) » (Lutz 2021)

Dans l'enquête, l'appartenance plus fréquente à la recherche quantitative qu'à la recherche qualitative est plus marquée pour les membres d'une communauté que pour les non-membres (Figure 12). Ceci dit, près de la moitié des répondants affirment que leurs travaux relèvent à la fois de la recherche qualitative et de la recherche quantitative.

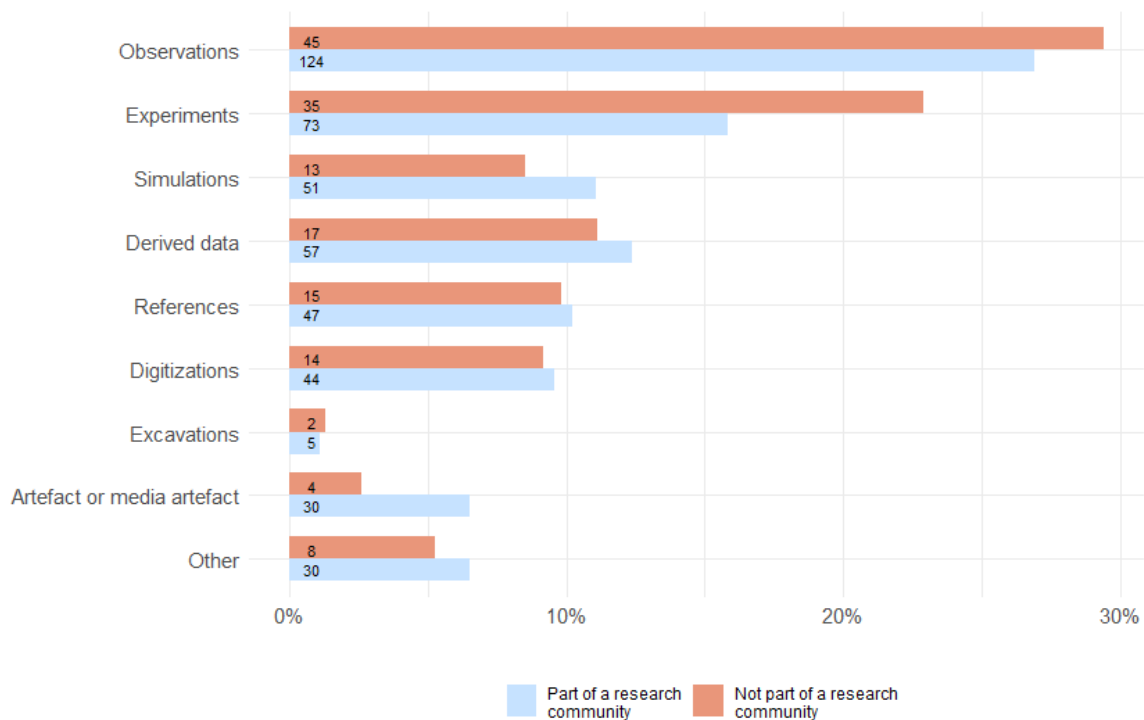
Figure 12 : Nature de la recherche



n= 240

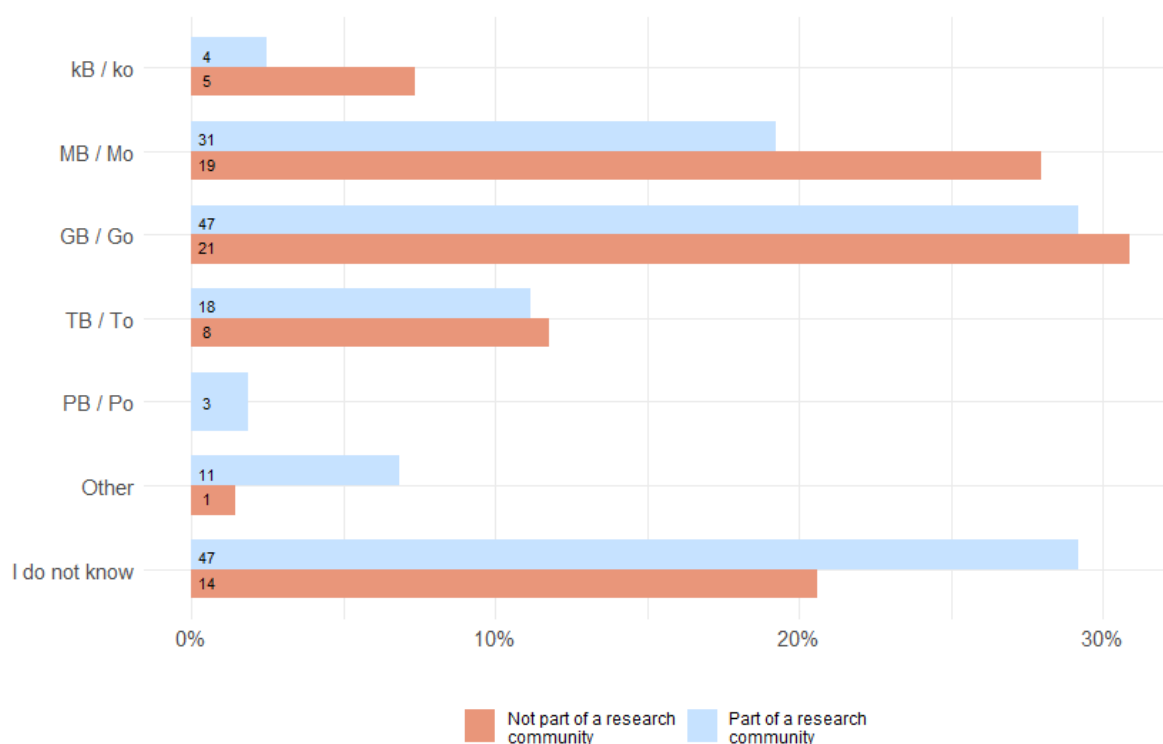
Deux questions de l'enquête sur le type de données manipulées et la taille typique des jeux de données permettent de caractériser plus en détail les données manipulées.

Figure 13 : Type de données



n= 614

Figure 14 : Taille des données



n= 229

En ce qui concerne le type de données, les plus citées par l'ensemble des chercheurs sont les données d'observation et les données résultant d'expériences (Figure 13). Les non-membres de communautés de recherche déclarent manipuler proportionnellement plus ces deux types de données que les membres de communautés. Les données de ceux-ci se répartissent en effet mieux dans les différentes catégories.

La taille typique des fichiers de données la plus fréquemment citée est le GB / Go et viennent ensuite le MB / Mo (proportionnellement surtout pour les non-membres), et la réponse « I do not know » dans une proportion proche des 30% pour les membres d'une communauté (Figure 14). Il est possible que cette proportion importante pour les membres de communautés vienne de la plus grande variété de la typologie de données, se traduisant également par une plus grande variété de la taille typique des fichiers, et donc par le choix plus fréquent de cette réponse.

Comme le montrent les résultats de l'enquête, la portée européenne ou internationale de la recherche, une échelle de temps typique des travaux ou des projets supérieurs à 5 ans et, dans une moindre mesure, le type de recherche plus quantitative sont notamment des caractéristiques qui semblent aller de pair avec le rattachement des chercheurs à une ou des communautés.

C. Focus sur deux profils particuliers : les communautés de données et les chercheurs « sans données »

1) Les communautés de données

Notre entretien avec Torsten Schwede et l'article de Berman et al. (2013) décrivent en détail le développement et le fonctionnement de la communauté « Protein Data Bank » (présenté et discuté ultérieurement). A partir de ces informations, un de nos objectifs était alors d'étudier si le concept de

« communauté de données » se retrouve ailleurs dans le paysage de la recherche académique suisse et de caractériser ses pratiques. Notre enquête cherche aussi à caractériser, à l'autre bout du spectre de cette typologie polymorphe des communautés, des chercheurs (et leurs pratiques individuelles) qui déclarent ne pas utiliser de données.

Sur les 230 chercheurs qui déclarent appartenir à une communauté de recherche, seulement 38 disent être dans une communauté de données. Outre cette faible représentation, la variété des réponses et certaines incohérences brouillent quelque peu la vision de ce type de communauté. De plus, comme beaucoup d'autres répondants, ces membres de communautés de données déclarent appartenir à plusieurs communautés, ce qui rend difficile une lecture linéaire des résultats. Un des éléments de pertinence choisi pour cette analyse est tout simplement la capacité à nommer clairement la ou les communautés évoquées et/ou à donner un lien vers une page web pertinente. Avec ces critères, on obtient alors moins de 20 réponses exploitables qui permettent d'identifier une vingtaine de potentielles communautés de données (Annexe 6).

Les trois grands domaines disciplinaires (sciences humaines et sciences sociales ; mathématiques, sciences naturelles et de l'ingénierie ; biologie et médecine) sont représentés. L'échelle de la recherche est au minimum nationale et pour la majorité internationale (11 chercheurs). Les données manipulées sont quantitatives ou quantitatives et qualitatives (mais pas uniquement qualitatives). On observe également qu'un quart environ de ces communautés est informel et un autre quart est lié à un projet. Ces éléments se veulent plus des indications pour d'éventuelles perspectives sur ces communautés de données, car le faible nombre de réponses et le flou présent dans la définition même de communautés de recherche au sein des chercheurs interrogés limitent la possibilité d'une caractérisation plus approfondie.

2) Les chercheurs « sans données »

A l'autre extrémité, lors des entretiens et de discussions préalables à ce mandat, des profils de chercheurs qui n'utiliseraient pas de données sont régulièrement apparus. Par exemple :

« Many researchers do not create data [...]. They use manuscripts, other kind of information than datasets that you can easily store. ». (Lutz 2021)

L'enquête visait donc à repérer les éventuelles pratiques de recherche sans aucune utilisation de données. Au final, seulement trois chercheurs déclarent ne pas utiliser de données dans l'enquête. Deux sont issus du domaine Economie – droit et le troisième de la musicologie. Deux d'entre eux déclarent par ailleurs suivre des pratiques individuelles de recherche. Deux de ces chercheurs déclarent publier en Open Access, démontrant ainsi qu'ils sont familiers et pratiquent au moins ce volet plus ancien que l'ORD de l'Open Science. D'après ces observations, le « mythe » du chercheur sans données semble ne pas correspondre à une large réalité, en dépit néanmoins de ce commentaire d'un chercheur en musicologie:

« Most of my research is philological/historical so it is expressed in texts and not in "data". »
(réponse à la question de l'enquête sous forme de texte libre)

Un autre aspect à considérer est que l'enquête, par son format numérique, n'a peut-être pas touché assez largement ce profil de chercheurs.

II. Rôle de l'organisation en « communautés de recherche » pour les pratiques ORD

Après avoir essayé de définir plus précisément les « communautés de recherche », l'objectif était de savoir si elles influencent sur le comportement et les pratiques des chercheurs en matière d'ORD et si oui de quelle manière. Pour cela, nous nous sommes intéressés à trois aspects :

- Est-ce que l'appartenance à une communauté peut atténuer les inquiétudes, minimiser les obstacles et augmenter les facteurs de motivation ressentis par les chercheurs ?
- Est-ce que l'appartenance à une communauté joue un rôle dans l'utilisation de standards par les chercheurs ?
- Est-ce que l'appartenance à une communauté peut être bénéfique au développement des compétences des chercheurs en ORD ?

A. Les facteurs de motivation et les obstacles pour les chercheurs au développement de l'ORD

Les entretiens et la revue de la littérature nous ont permis de repérer divers éléments qui pouvaient interférer positivement ou négativement dans le développement des pratiques ORD. Ces éléments ont été répartis en trois groupes :

- Les inquiétudes (« concern ») pour les éléments moins tangibles liés à des craintes plus subjectives ;
- Les obstacles pour les éléments tangibles gênant directement le partage de données ;
- Les facteurs de motivation (« incentives ») pour les éléments encourageant (voire obligeant) les chercheurs à partager leurs données.

Comme le souligne Torsten Schwede :

« One of the precondition for OS to work is that we have understanding for the motivations and the limitations of all players in the communities. Otherwise, there is no motivation to share anything » (Ackermann, Schwede 2021)

Il semble effectivement essentiel de prendre en compte ces facteurs pour mieux comprendre comment agir dessus.

1) Les préoccupations ressenties par les chercheurs dans les pratiques ORD

Les personnes interrogées durant les entretiens ont souvent mentionné des « craintes » qui semblaient gêner le développement des pratiques ORD des chercheurs. La synthèse de ces craintes évoquées dans les entretiens a permis de construire une série de propositions de réponses pour la partie de l'enquête sur les inquiétudes. Le Tableau 3 illustre ces propositions à l'aide d'exemples issus des entretiens.

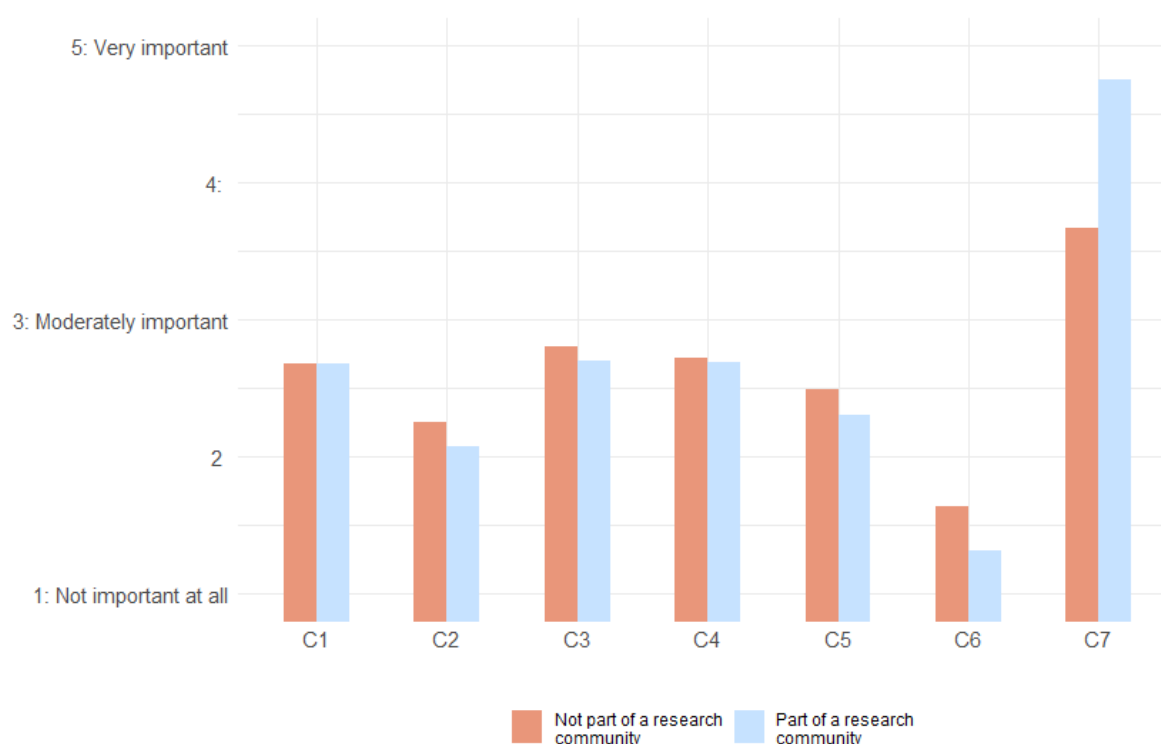
Tableau 3 : Liste des inquiétudes avec extraits des entretiens

Extraits des entretiens	Proposition de réponse dans l'enquête
« Craintes sur la réutilisation des données et le risque que ses données soient mal utilisées » (Burgi 2021)	I am concerned about my data being not well handled and used.

« They want time for their publications ; make sure they can use the data their own way ; to keep control on their data » (Gautschy, Subotic 2021)	I feel that I might lose control on my data I am concerned not to be properly acknowledged as the data producer.
« In the medical domain (personalized health data network): if "I share my data": • It would take something away from me • people may see errors in it and that would be embarrassing for me • people may reap benefits from my data without me getting anything out of it » (Ackermann, Schwede 2021)	I am concerned that someone else publishes results from my data before I do.
« L'acquisition des données représente le gros du travail scientifique et les donner gratuitement serait équivalent à travailler pour rien » (Burgi 2021)	I think that the benefit would be too low as compared to the efforts.
« Si les données peuvent être facilement produites, il y a moins de motivation à les partager. » (Burgi 2021)	I do not see the point.

Les chercheurs pouvaient évaluer le degré d'importance de leurs réponses sur une échelle de 1 à 5.

Figure 15 : Inquiétudes concernant les pratiques ORD



n= 163

- C1** : I am concerned about my data being not well handled and used
C2 : I feel that I might lose control on my data
C3 : I am concerned that someone else publishes results from my data before I do
C4 : I am concerned not to be properly acknowledged as the data producer
C5 : I think that the benefit would be too low as compared to the efforts
C6 : I do not see the point
C7 : Other

La Figure 15 montre qu'aucun des éléments mentionnés n'est jugé comme réellement important cependant ils apparaissent tous comme des inquiétudes existantes. Seul le C6 « I do not see the point » semble moins pertinent. La catégorie C7 « Other » était laissée en texte libre pour laisser la possibilité aux chercheurs d'indiquer des inquiétudes qui n'apparaîtraient pas dans la liste. Logiquement, les chercheurs qui prennent le temps de compléter ce texte libre jugent leur inquiétude comme très importante. En tout, douze propositions complémentaires ont été ajoutées, elles peuvent être regroupées sous cinq grands points :

- Problématique liée aux particularités des données qui rendent impossible le partage
« *Context information is crucial for ethnographic projects; if data is anonymised/masked; it loses its potential use value both for me and further researchers.* ».
- Problématique d'anonymisation des données notamment en santé
- Manque de soutien dans la gestion des données
« *Main concern: how to make sensitive data accessible in an appropriate way without having enough time and financial or technical support from the university.* ».
- Problématique technique
« *My concerns are mostly technical: current ORD platforms are too primitive for my needs.* ».
- Problématique de temps

Les propositions qui émanent de ces réponses sont très factuelles et font écho aux propositions d'obstacles listées dans la question suivante. Les inquiétudes bien qu'existantes ne semblent pas avoir des effets très marqués sur les pratiques ORD. Surtout, on ne remarque pas de différence significative entre les inquiétudes des chercheurs dans et hors communautés.

2) Les obstacles au développement de l'ORD

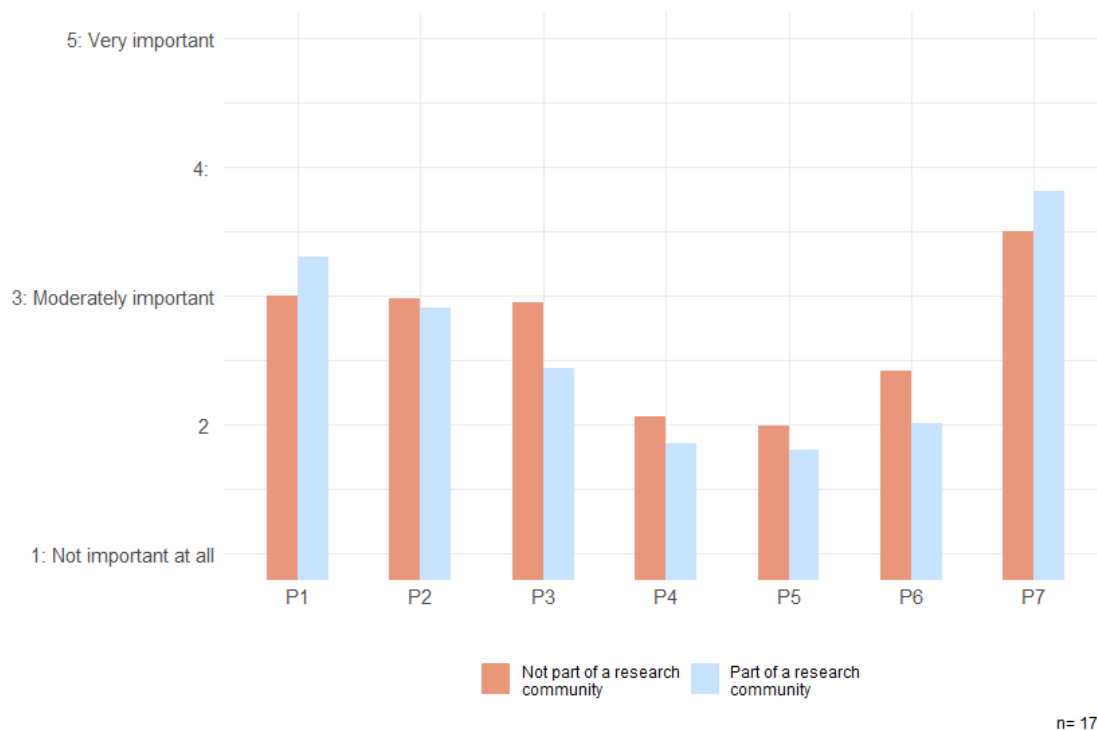
Sur le même principe que les inquiétudes, une liste des obstacles a été établie à la suite de la revue de la littérature et des entretiens (Tableau 4).

Tableau 4 : Liste des obstacles avec extraits des entretiens

Extraits des entretiens	Proposition de réponse dans l'enquête
« As soon as they are individual data it gets very tricky (in education sciences, data collected from kindergartens). Video tapes: almost impossible to anonymize. How to make available sensitive data and secure access to data and documentation? » (Lutz 2021)	There are restricted conditions on the use of the data that I handle (and thus on sharing them).
« Medical domain: not much time to do research. They are afraid that if they make their data available, scientists with more time for research just steal their glory and treat them as someone delivering samples and observations. » (Ackermann, Schwede 2021)	I do not have enough time to prepare the data and to make them accessible and reusable.
« If the time scale is long, people are less willing to share their data, they wait for the project to be finished. » (Lutz 2021)	The typical time scale of my research is very long (and I will not share my data before the project is finished).
« Sometimes data are extremely complex, and preparing or arranging them for re-use takes some skills. They miss the knowledge, the skills » (Gautschy, Subotic 2021)	I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable.
« Magnitude of the data. Some disciplines produce terabytes of data. How to handle them? » (Ackermann, Schwede 2021)	I think that no appropriate technical tool is available for sharing my data and making them reusable.
« People need more skills and a better education at the individual scale. The education part is the most important. You can always offer more, but without education, providing more services would be useless. » (Gautschy, Subotic 2021)	I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable.

On retrouve dans ce tableau les éléments plus factuels qui avaient été rajoutés dans les inquiétudes. L'importance de ces éléments a été jugée plus conséquente dans les résultats de l'enquête comme le montre la Figure 16.

Figure 16 : Obstacles pratiques concernant le partage et la réutilisation des données



P1 : There are restricted conditions on the use of the data that I handle (and thus on sharing them)

P2 : I do not have enough time to prepare the data and to make them accessible and reusable

P3 : The typical time scale of my research is very long (and I will not share my data before the project is finished)

P4 : I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable

P5 : I think that no appropriate technical tool is available for sharing my data and making them reusable

P6 : I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable

P7 : Other

Les deux obstacles évalués comme les moins importants sont la question des compétences et l'accès à des outils adaptés (Figure 16). L'obstacle des restrictions sur l'accès aux données est majoritaire. Ces restrictions ont été détaillées dans la suite de la question et elles sont majoritairement légales, dans une moindre mesure éthiques et, pour quelques chercheurs, commerciales.

Les obstacles répertoriés dans le texte libre de la case *Other* sont plus nombreux que ceux liés aux inquiétudes. En tout, une trentaine de réponses ont été complétées, elles peuvent être regroupées autour de ces cinq thématiques :

- Manque d'un soutien institutionnel

« There is no institutional awareness about providing infrastructure for sustainability of research results not just for simple preservation. »

- Crainte qu'il n'y ait des erreurs dans les données
- Réticence d'autres personnes impliquées dans la recherche (collègues, superviseur, participants...)

« Participants (e.g., journalists, communicators, strategists) would most likely not participate in our studies, if they knew the data was published. »

- Manque d'intérêt pour le partage et surtout absence d'incitation financière
« I don't care much about sharing my data (at least for free). »
- Manque d'intérêt des données
« I am not convinced that my qualitative data may be of interest for other researchers and that they can be re-used. »

Toutes ces thématiques apparaissent une ou deux fois, seule la question financière est citée trois fois. La prise en compte de ces obstacles semble incontournable pour améliorer les pratiques en ORD surtout qu'ils touchent l'ensemble des chercheurs sans distinction significative entre chercheurs dans et hors communauté.

3) Les facteurs de motivation au développement de l'ORD

Les facteurs de motivation font écho aux obstacles dans le sens où le fait de les lever peut être un facteur de motivation. On peut prendre comme exemple l'obstacle financier identifié dans les propositions des chercheurs qui peut devenir un facteur de motivation comme dans les exemples suivants (Tableau 5).

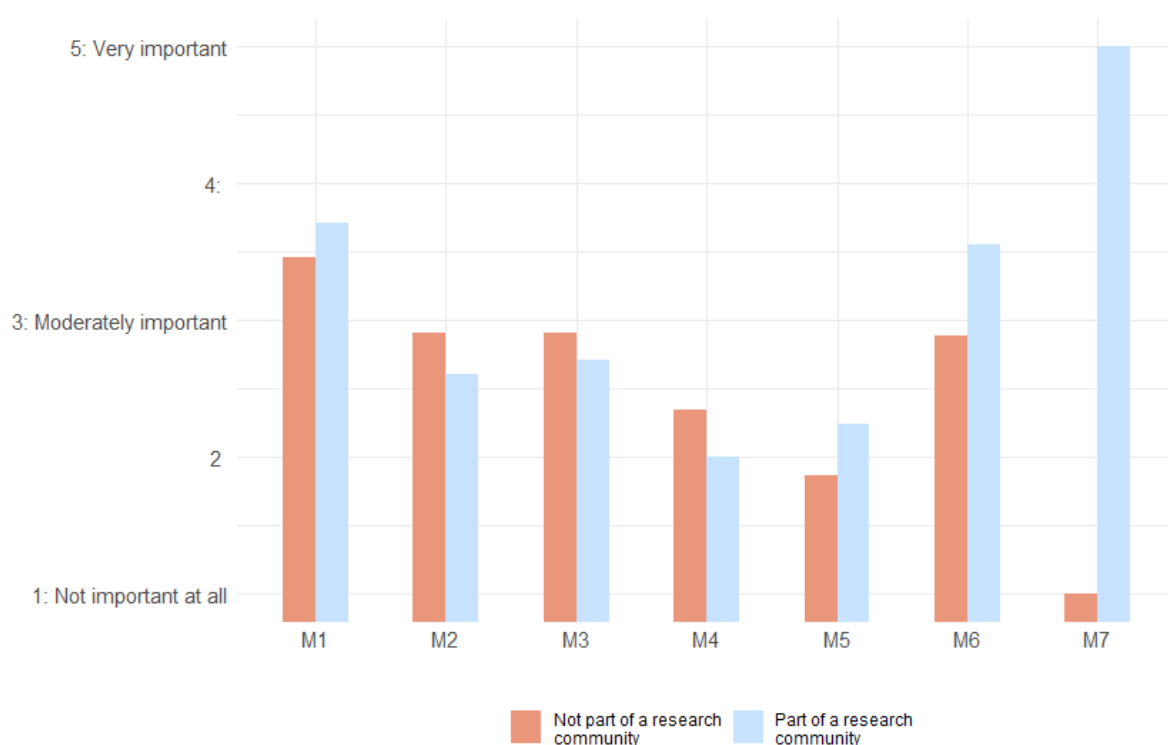
Tableau 5 : Liste des facteurs de motivation avec extraits des entretiens

Extraits des entretiens	Proposition de réponse dans l'enquête
<i>« There is a change of culture, with a push from the funders and the journals (ask for replication material). current "stick" approach by funders and journals (ex.: your next project will be funded only if you shared your data from the current one). » (Lutz 2021)</i>	It complies with the requirements from (funding agencies or) publishers.
<i>« There is an understanding in the communities that you also have to deposit the raw data just to prove that you are doing high quality science. If you want some credibility for your field, you have to play it open. » (Ackermann, Schwede 2021)</i>	It brings me academic recognition (evaluation criterion for my academic career).
<i>« We do not do it because FNS tells us to, but I want to do it because I get benefit from it, personally or indirectly (people see my research, or I get more visibility, ...). » (Ackermann, Schwede 2021)</i>	It brings me scientific recognition.
<i>« What is really missing is the understanding by the funding agencies that once a community comes together with an agreement for standards and archive,</i>	I would like to benefit from the financial incentives existing for encouraging ORD practices.

<i>they need to finance this archive [to support] the effort driven by the community for the high quality curation of the data, to make them admit that this kind of service adds value to the community. »</i>	
« ...an agreement in the community that you cannot publish a paper if you do not deposit your data into that archive. » (Ackermann, Schwede 2021)	It is very much practiced in my community.
« You identify people who are open-minded and willing to engage, from a good distribution of universities (including universities of applied sciences) and of disciplines. » (Lutz 2021)	I commit to ORD because it corresponds to a persistent belief.

Les résultats de l'enquête sur ces facteurs de motivation sont plus variés (Figure 17).

Figure 17 : Facteurs de motivation pour les pratiques ORD



n= 169

- M1:** It complies with the requirements from funding agencies or publishers
M2: It brings me academic recognition (evaluation criterion for my academic career)
M3: It brings me scientific recognition
M4: I would like to benefit from the financial incentives existing for encouraging ORD practices
M5: It is very much practiced in my community
M6: I commit to ORD because it corresponds to a persistent belief
M7: Other

L'influence des bailleurs de fonds sur la motivation (et on retrouve ici la question financière) apparaît comme importante. A un niveau d'importance similaire, on note également l'idée d'une forme de conviction profonde "persistent belief" qui pourrait avoir un lien avec la notion de « culture de partage ». La partie "Other" des motivations est la plus marquée avec une évaluation à un niveau "very important". En réalité, seulement huit propositions ont été ajoutées mais toutes évaluées comme essentielle :

- Solidarité entre chercheurs
« *To make data accessible to researchers with no/less funding* »
- Souhait de faire avancer la science et la société (cité quatre fois)
- Utilisation de fonds public (cité deux fois)
« *Public money is used to produce such data and they should be exploited / I would never be able to exploit the data at the 100% of their potential myself not will my research team* »
- Suivi des recommandations de l'UNESCO pour l'OS

Les motivations ajoutées sont pour la plupart liées à des valeurs sociétales et font écho aux convictions personnelles qui apparaissent aussi comme des moteurs pour le développement des pratiques ORD. Les résultats de la Figure 17 montrent que ces convictions sont plus importantes chez les chercheurs dans les communautés de recherche que pour ceux hors. Cependant, comme pour les inquiétudes et les obstacles, cette différence n'est pas significative.

L'ensemble des facteurs de motivation évoqués sont à comprendre et à activer dans le cadre d'une stratégie ORD. Dans ce sens, Beate Böckem apporte une vision plus pragmatique qui relève la pertinence d'une approche bottom-up plus efficace qu'une approche top-down :

« *Our demand, and one of the incentives of our researchers, is to not having to change ourselves to fit in, but to add to the field of data management.* » (Böckem, Schumacher 2021)

Les inquiétudes, les obstacles et les facteurs de motivation sont des éléments à prendre en compte. On ne constate pas de réelle différence entre les chercheurs dans et hors communauté. Il émane plutôt une forme de consensus sur ces sujets avec des degrés d'importance similaires.

B. L'utilisation de standards

La question de la standardisation a été très présente dans les entretiens, avec une opinion partagée quant à la nécessité d'aller vers un maximum de standardisation.

« *Standardization is crucial. Need of at least one better data scheme per discipline, acceptance of how things should be documented and described, in a disciplinary way. "It is where the community-building aspect comes into place. Not wise to impose standards. Create a frame, a structure where the dialog happens.* » (Lutz 2021)

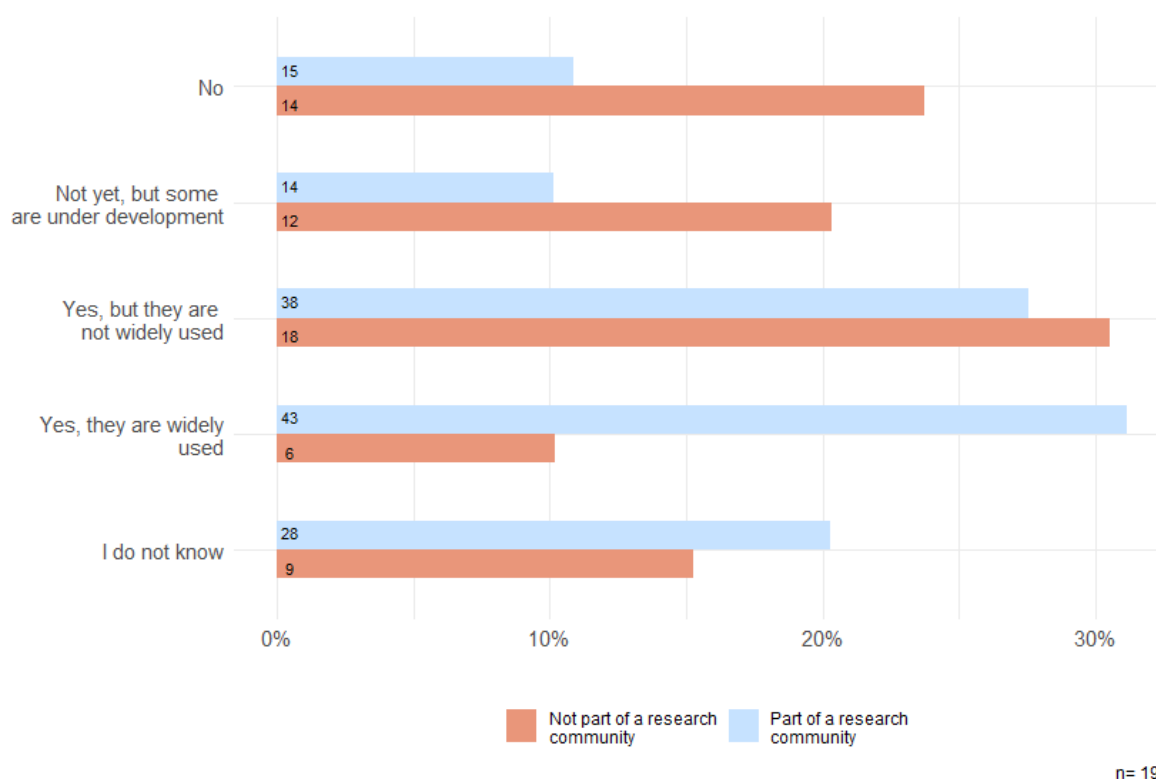
« *Unless the data is based on common standards and understanding (on how to annotate the data, the metadata, formatting, interoperability,...), it is a waste of disk space - Community driven semantic standardization type of efforts.* » (Ackermann, Schwede 2021)

« *[C'est aux communautés] d'amener les standards. Sans contrainte forte de standardisation, il est difficile d'envisager cette réutilisation.* » (Burgi 2021)

Pour Pierre-Yves Burgi, si la standardisation est nécessaire, les standards doivent être amenés par les communautés pour correspondre complètement aux réalités du terrain.

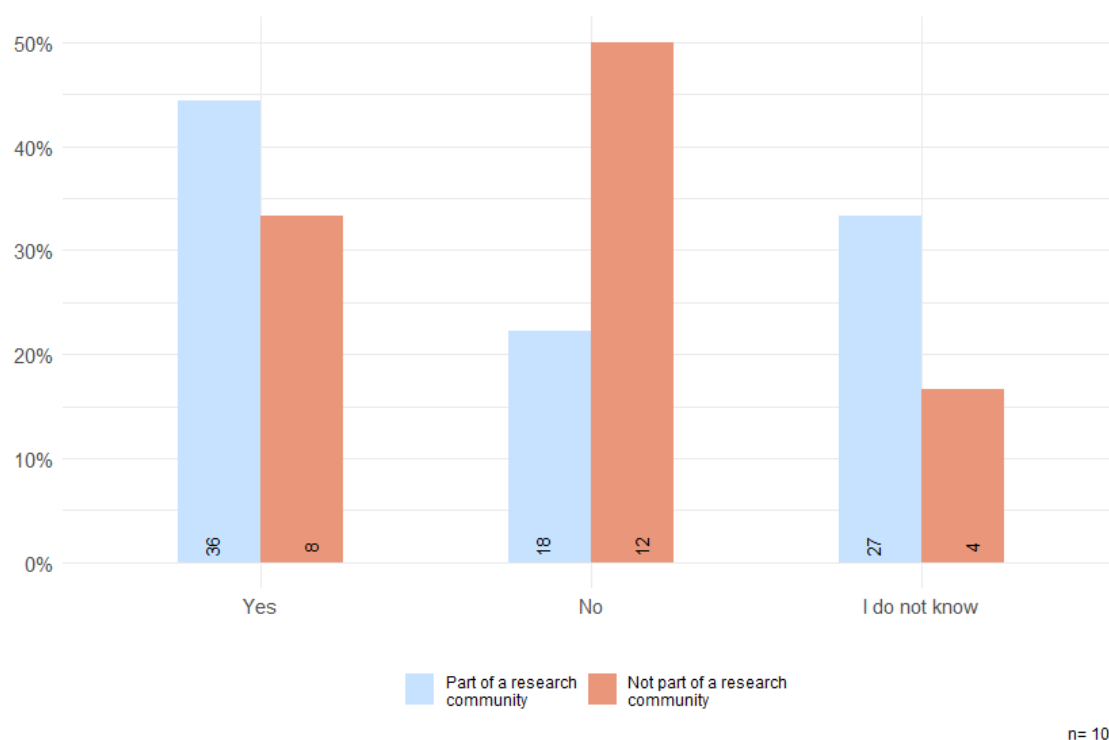
La nécessité de la standardisation est donc bien présente dans les entretiens. Par contre, elle est plus difficile à décélérer à partir des résultats de l'enquête auprès des chercheurs.

Figure 18 : L'utilisation de standards pour les données



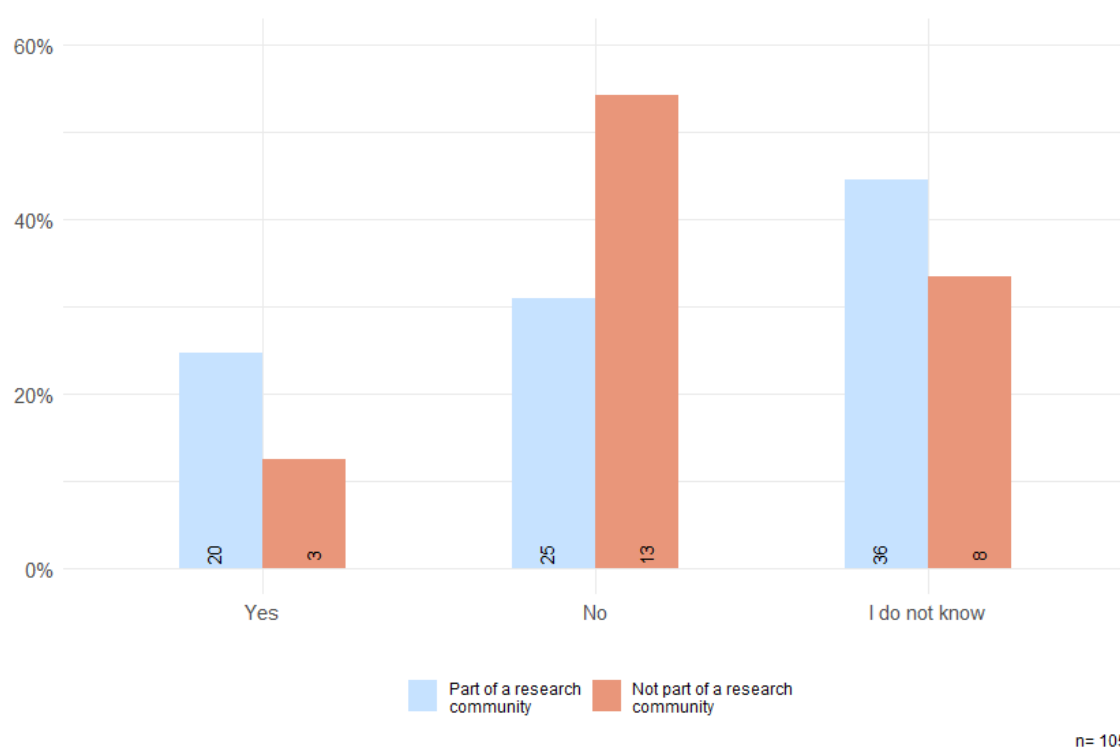
On constate que proportionnellement les standards pour les données en général sont nettement plus utilisés couramment par les chercheurs qui sont membres de communautés que par les chercheurs non membres (Figure 18). Cependant, c'est également ces profils qui déclarent en proportion assez importante (autour de 20%) « ne pas savoir » si des standards existent dans leur communauté. Parmi les non-membres de la communauté, il faut noter qu'une proportion significative des réponses indique qu'ils ont connaissance de standards en cours de développement ou existants mais sans qu'ils soient encore utilisés largement.

Figure 19 : L'utilisation de standards pour les métadonnées



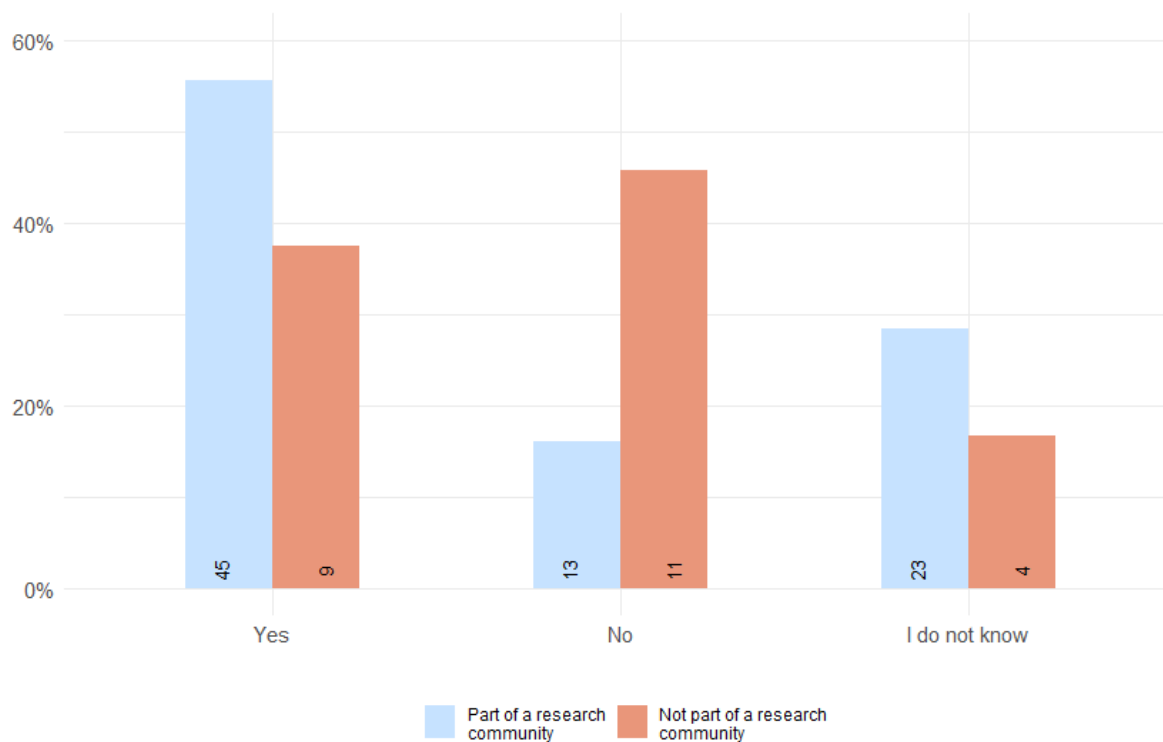
Plus en particulier, les standards sur les métadonnées sont proportionnellement plus utilisés par les membres de communautés, bien qu'environ un tiers de cette catégorie déclare aussi ne pas savoir s'ils sont utilisés au sein de la communauté (Figure 19). A noter que les non-membres, s'ils répondent en majorité qu'ils n'utilisent pas ce type de standards, répondent aussi en nombre (environ 30% d'entre eux) qu'ils les utilisent.

Figure 20 : L'utilisation de « PIDs »



Les PIDs ne sont pas utilisés aussi couramment (Figure 20). Près de la moitié des non-membres de communautés déclarent ne pas les utiliser et dans les 2 catégories, une proportion importante des réponses dit ne pas savoir.

Figure 21 : L'utilisation de standards de format de fichiers

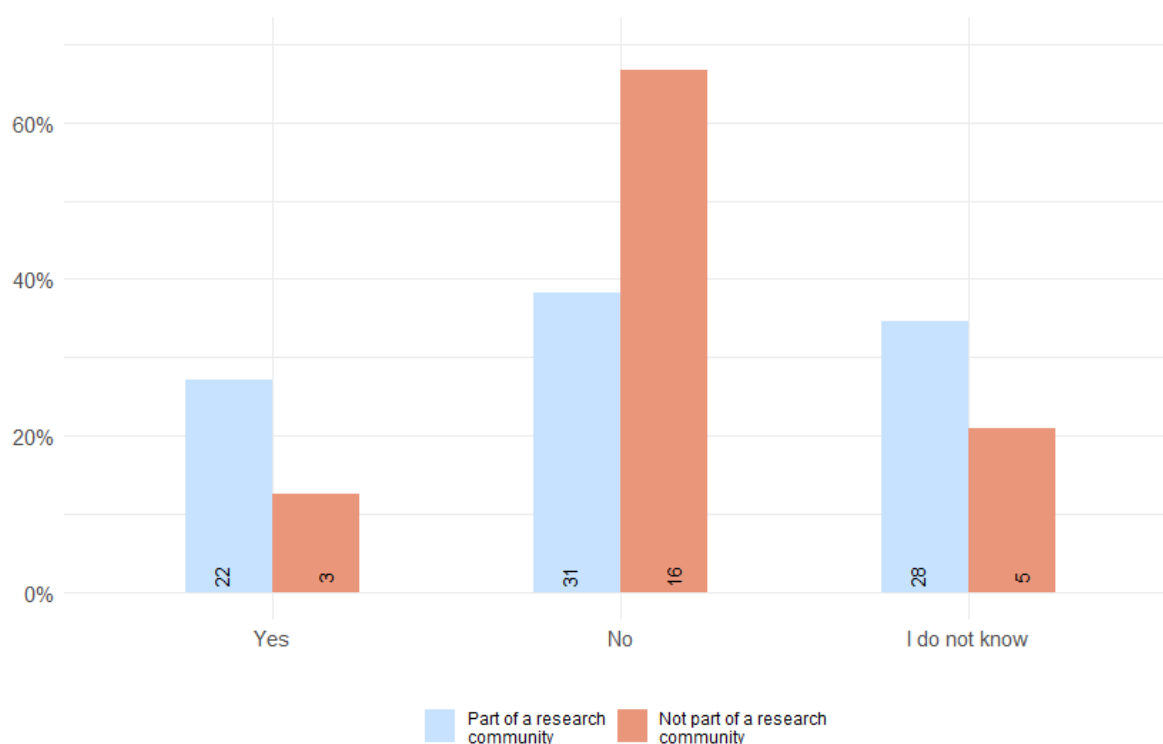


n= 105

L'utilisation de standards sur les formats de fichiers est bien marquée chez les membres de communautés (un peu plus de la moitié des réponses dans cette catégorie). La plus grande partie des réponses des non-membres se répartit entre utilisation et non-utilisation, avec une légère prédominance pour cette dernière (Figure 21).

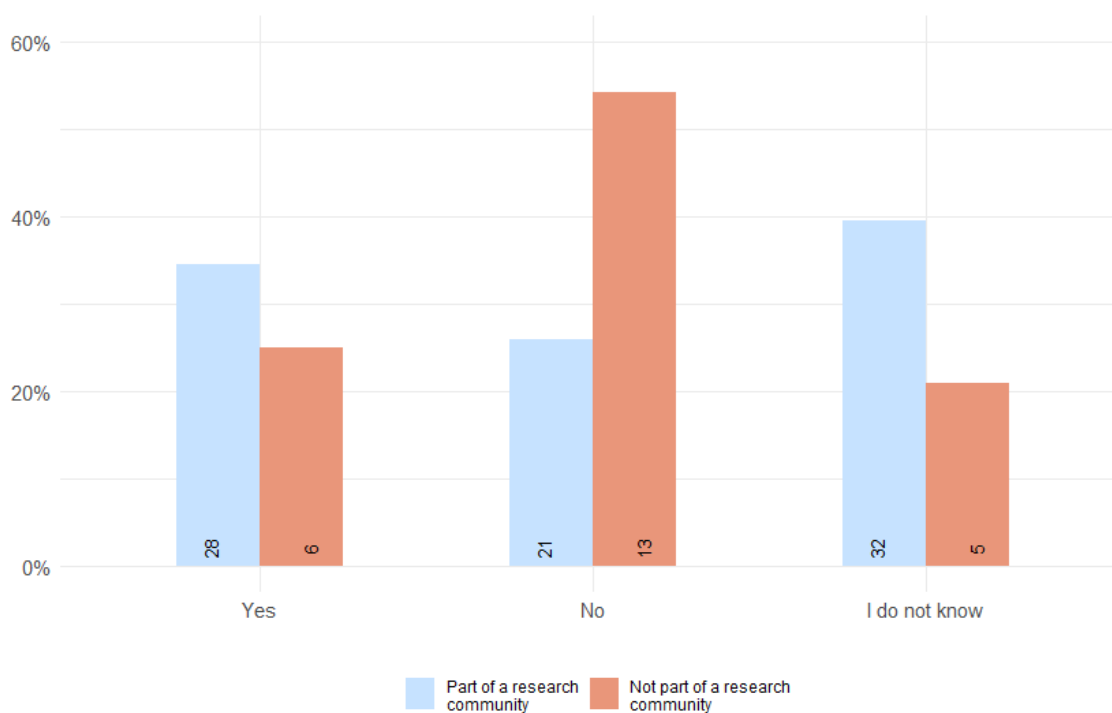
Les réponses sur l'utilisation de standards sur la documentation sont majoritairement négatives pour les non-membres mais également (bien que de façon moins marquée) pour les membres de communauté (Figure 22). On peut expliquer en partie ce résultat par la possible ambiguïté du terme « documentation standards » utilisé dans cette question : standards sur les métadonnées (abordés avec une question précédente) ou standards sur les formats de fichiers texte qui documentent les données (aussi abordés avec une autre question).

Figure 22 : L'utilisation de standards pour la documentation



n= 105

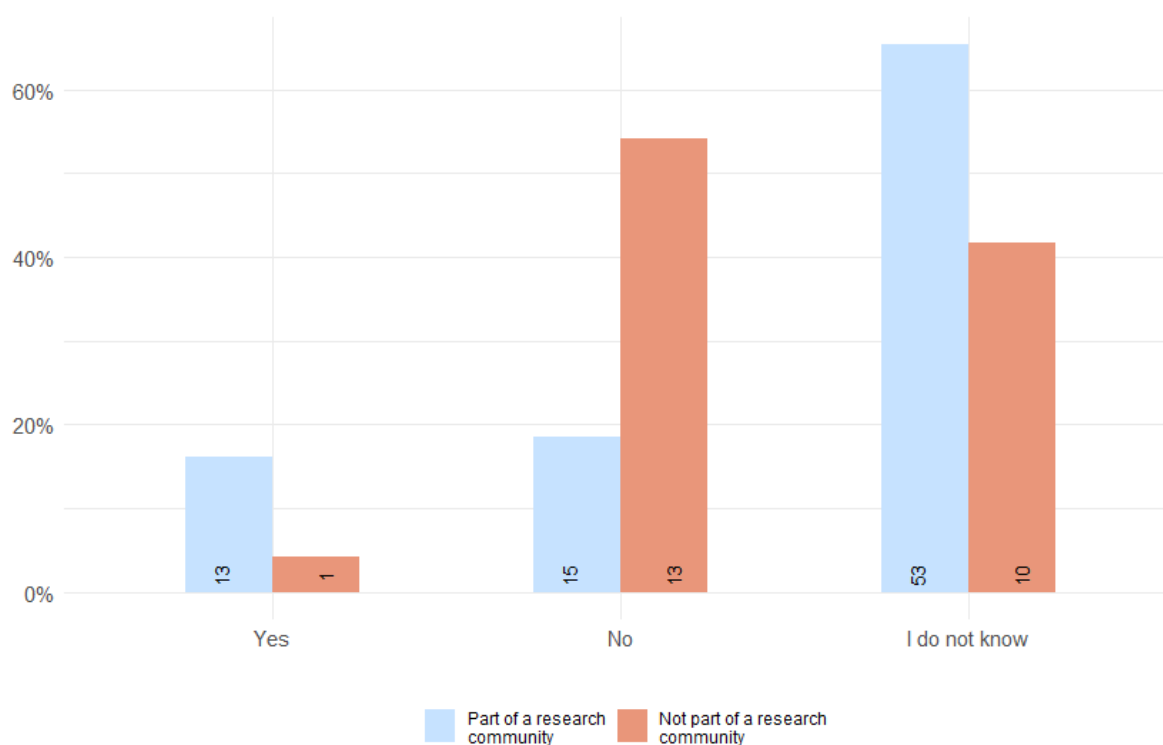
Figure 23 : L'utilisation de standards pour les licences



n= 105

Concernant les standards sur les licences, les non-membres déclarent en majorité ne pas en utiliser (Figure 23). En revanche, les membres de communautés déclarent le plus souvent ne pas savoir s'ils sont utilisés dans leur communauté. Vient ensuite le nombre de réponses « oui », puis (mais dans une proportion assez proche) le nombre de réponses « non ».

Figure 24 : L'utilisation d'autres standards



n= 105

Enfin, concernant d'autres standards que ceux proposés précédemment, les membres de communautés déclarent en majorité qu'ils ne savent pas si d'autres standards sont utilisés au sein de la communauté (avec une proportion minoritaire de réponses « oui »). Une majorité de non membres répond qu'elle n'utilise pas d'autre standard (Figure 24).

Pour résumer cette partie sur la connaissance et l'utilisation des standards, on peut donc dire que les membres de communautés utilisent plus couramment des standards que les non-membres de communautés. Ces résultats sont confirmés en particulier pour les standards sur les métadonnées et ceux sur les formats de fichiers. Une proportion importante (entre 20% et plus de 40% selon les questions) de membres de communautés déclare pour plusieurs de ces questions ne pas savoir si les standards (en général ou certains standards particuliers) sont utilisés au sein de leur communauté.

C. Le niveau de compétences en ORD

Durant les entretiens, le cycle de gestion des données a été utilisé comme base de discussion autour des pratiques en ORD et de leur degré d'avancement. Pour l'enquête, visant directement les chercheurs, il était plus pertinent d'aborder cet aspect à partir d'éléments et de tâches pratiques auxquels les chercheurs pouvaient répondre sans avoir forcément une connaissance théorique poussée sur les approches en gestion des données de la recherche.

Les activités abordées étaient les suivantes :

- Producing a Data Management Plan (DMP)

- Preparing your data for uploading them into a repository
- Uploading research data into a repository
- Sharing research data via a repository
- Searching research data produced by others
- Re-using research data produced by others
- Reproducing research results produced by others

Pour chaque activité, le même éventail de choix de réponses était proposé, dans le but d'évaluer le niveau de compétences et le degré d'autonomie du chercheur pour mener à bien la tâche.

- I am able to do it alone
- I do not know if I am able to do it
- I am able to do it with help from a librarian
- I am able to do it with help from my community
- I am able to do it with help from my institution
- I am able to do it with help from the data infrastructure I use
- I am aware of the process but I do not know how to do it

Pour les sept tâches proposées, la plus grande partie des membres de communautés est capable de réaliser la tâche de façon autonome (voir Figure 25 à Figure 31). Le reste des réponses se répartit de façon variable. L'aide de la communauté, de l'institution, ou de l'infrastructure de données utilisées jouent un rôle dans plusieurs cas. Le soutien de la communauté est particulièrement important pour découvrir, réutiliser des données et reproduire des résultats de recherche. En revanche, les résultats montrent que pour toutes les tâches, le soutien apporté par les bibliothécaires est plus faible que le soutien apporté par d'autres moyens ou personnes.

La proportion des non-membres de communautés capables de réaliser une tâche de façon autonome est également importante. Comme pour les membres, elle est majoritaire pour toutes les tâches, sauf en ce qui concerne la production d'un DMP et la reproduction de résultats de recherche, et, dans une moindre mesure, le partage des données sur un dépôt. Dans ces deux cas, la majorité des chercheurs disent ne pas savoir s'ils sont capables d'effectuer cette tâche.

L'analyse des résultats pour ces questions a ensuite été restreinte aux réponses des membres de communautés de données seulement. Elle démontre un comportement assez similaire à celui des membres de communauté tel que discuté ci-dessus.

Figure 25 : Élaboration d'un Data Management Plan (DMP)

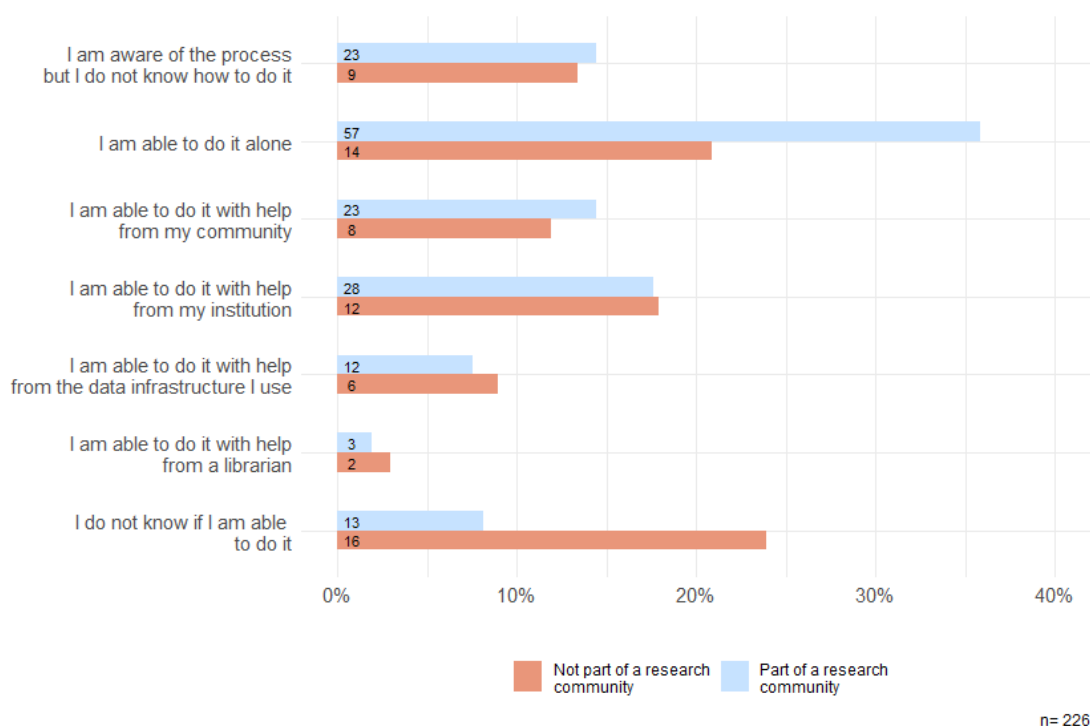


Figure 26 : Préparation des données pour les télécharger sur un dépôt

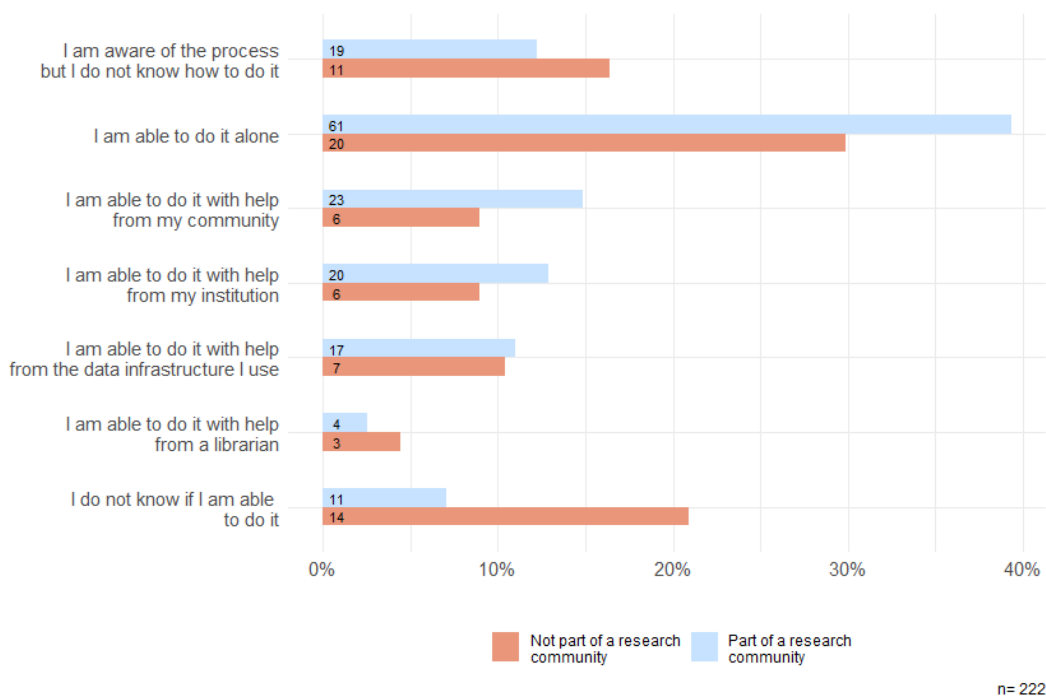
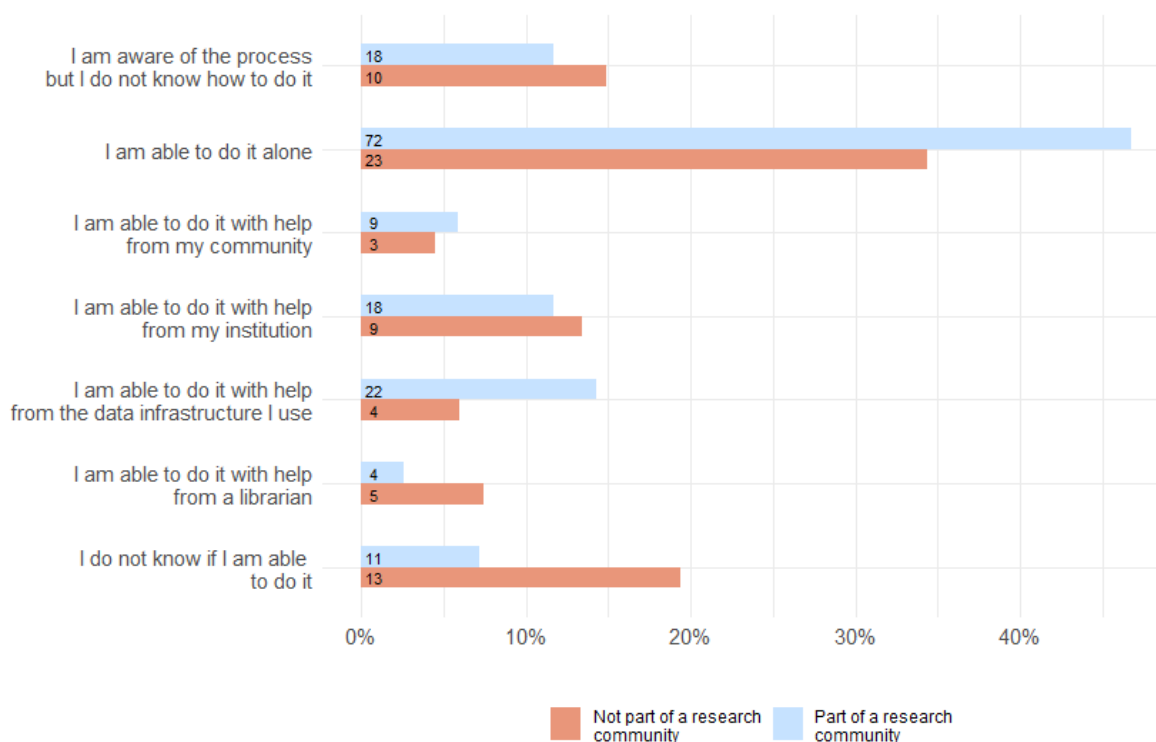
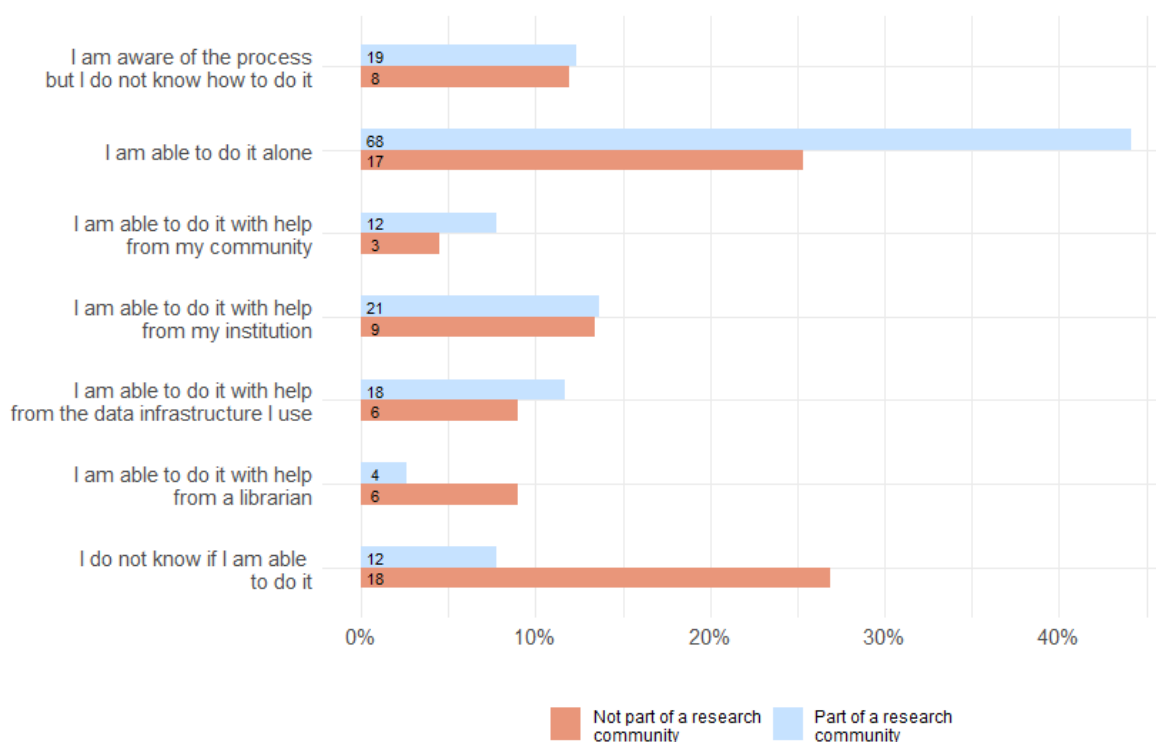


Figure 27 : Téléchargement des données de recherche sur un dépôt



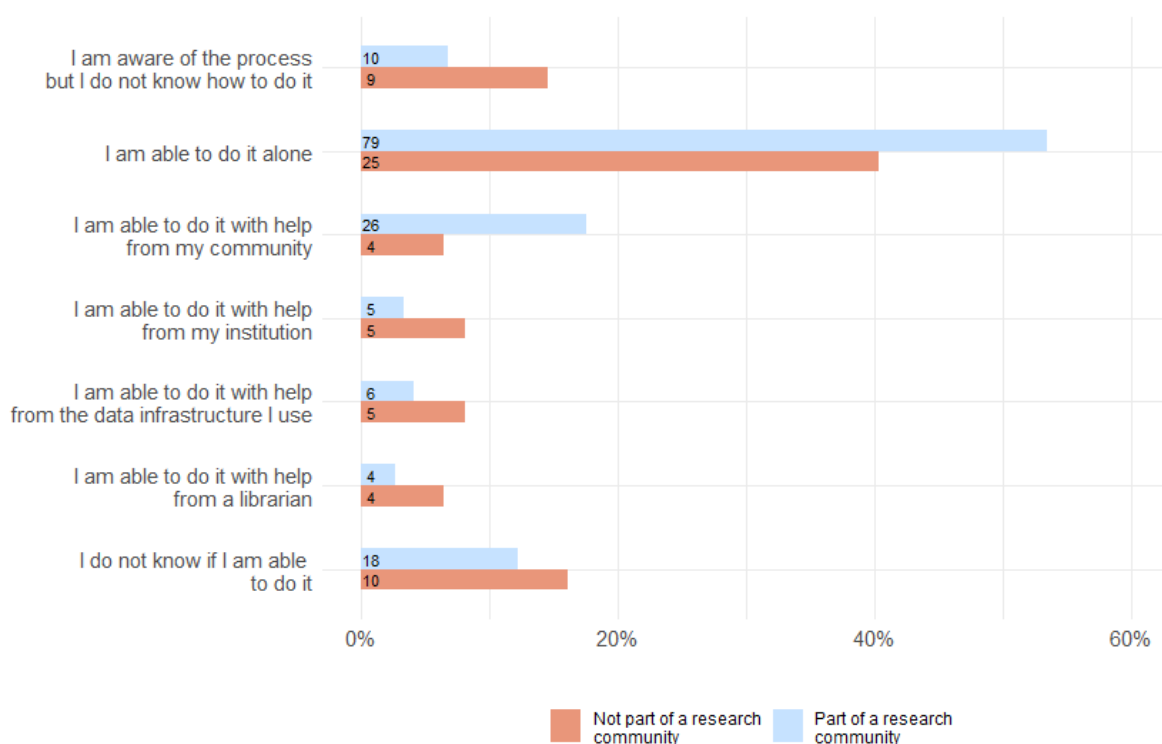
n= 221

Figure 28 : Partage des données de recherche via un dépôt



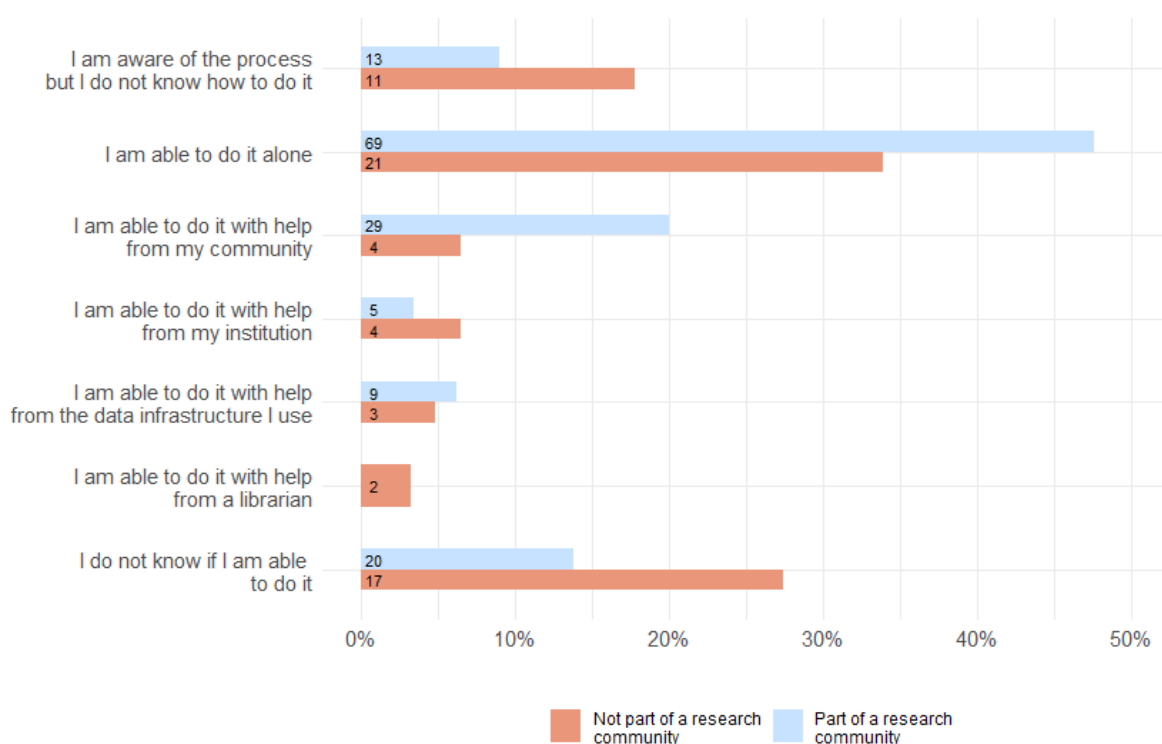
n= 221

Figure 29 : Recherche de données de recherche produites par d'autres



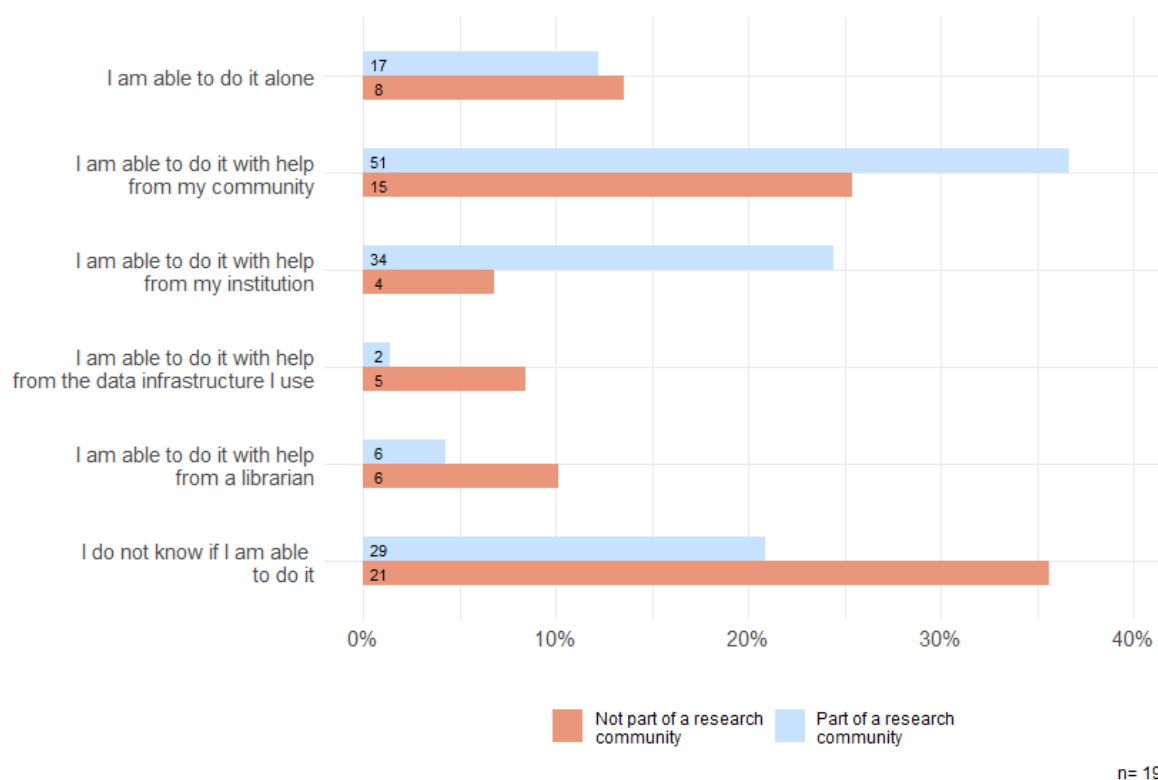
n= 210

Figure 30 : Réutilisation des données de recherche produites par d'autres



n= 207

Figure 31 : Reproduction de résultats de recherche produits par d'autres



En résumé de cette partie, on peut dire que, d'après ces résultats, le fait d'appartenir à une communauté est corrélé avec un plus grand degré d'autonomie pour réaliser les tâches proposées. Les non-membres peuvent aussi montrer un certain degré d'autonomie, mais dans une proportion plus faible et pour moins de tâches. Le soutien apporté par les bibliothécaires est insignifiant, au contraire de celui apporté par la communauté, l'institution, ou l'infrastructure de données.

On constate que l'appartenance à une communauté est corrélée avec l'utilisation de standards et des compétences plus développées en ORD. Ces deux éléments semblent être des bons candidats pour l'évaluation du niveau de maturité des pratiques ORD par une communauté. Ce sont aussi des leviers potentiels pour le perfectionnement des pratiques en ORD pour chaque communauté.

III. Les leviers possibles pour le développement de l'ORD

A. Acteurs-clés

L'enquête incluait aussi une question avec réponse en texte libre sur les acteurs possibles pour améliorer les pratiques en ORD. Les réponses sont variées. Le Tableau 6 les présente par catégorie, avec le cas échéant des détails sur les modes d'implication possibles de ces acteurs.

Tableau 6 : Acteurs-clés pour l'ORD

Catégorie et sous-catégories	Action proposée pour les impliquer	Contribution possible
Cadre institutionnel		
Collègues chercheurs, chercheurs seniors	♦ Les sensibiliser	
Chercheurs des disciplines qualitatives (par exemple, des anthropologistes, humanités)	♦ Prendre en compte leurs besoins différents ; les impliquer dans la discussion sur les approches, les solutions et les perspectives de leur point de vue	
Professeurs, enseignants	♦ Les sensibiliser	
Pls		♦ Promouvoir les valeurs ORD lors de la formation des jeunes chercheurs.
Tous les responsables institutionnels (recteurs, directeurs, etc.) et chefs de département, de laboratoire ou de groupes de recherche	♦ Stimuler leur engagement	♦ Entraîner leurs collaborateurs dans la dynamique.
Etudiants	♦ Les sensibiliser le plus tôt possible	
Universités et institutions de recherche au niveau du leadership		♦ S'engager plus, fournir plus de RH pour soutien – avec des personnes désignées ♦ Fournir de meilleurs outils pour la gestion des données au niveau du groupe de recherche (dépôt de données), par exemple, des solutions institutionnelles radicalement différentes que ce celles proposées actuellement par les bibliothèques (et permettant des requêtes dynamiques pour la découvrabilité) ♦ Fournir un soutien sous forme d'allocation de financement et de temps supplémentaire, de support technique et de clarifications sur les aspects légaux. ♦ Reconnaître et récompenser le

		partage des données – en allouant des financements dédiés, en particulier aux grands projets empiriques terminés. ♦ Prendre en compte la contribution des chercheurs pour l'ORD dans les procédures d'engagement, d'évaluation et d'avancement de leur carrière (suivre les principes DORA ⁷).
Bureaux de transfert technologique		
Services IT, experts données et IT		♦ Fournir des outils et des portails conformes aux plus hauts standards de qualité
Bibliothécaires		
Agences de financement et faïtières de la recherche		
Autorités fédérales (financement et planification à long terme); SNSF, CUSO; agences gouvernementales, départements fédéraux rectorats		♦ Financer les data managers pour leur contribution à l'ORD ♦ Imposer des contraintes ♦ Fournir des dépôts de données sûrs et centralisés
Organismes d'accréditation des universités		♦ Contraindre les universités à implémenter les politiques ORD ♦ Conditionner l'attribution de fonds à la production d'un DMP conforme aux principes ORD ♦ Développer des politiques plus adaptées aux sciences humaines et sociales
Dépôts	♦ Impliquer le personnel des infrastructures de recherche	♦ Améliorer et simplifier les outils pour la soumission des données et la curation des métadonnées
Editeurs (En particulier ceux des plus grandes revues scientifiques (Elsevier...))		
Archivistes		
Musées		
Gestionnaires de l'information		
Développeurs		♦ Développer des méthodes de partage efficaces pour le partage des données

Des éléments d'information plus détaillés et contextualisés sont disponibles dans les entretiens et discuté ci-dessous.

Communautés de recherche

Les chercheurs doivent être impliqués, par exemple aux étapes-clés de la gestion de leurs données et de leur partage (sur l'infrastructure DaSCH): « *it helps if the researchers are part of the whole process,*

⁷ Déclaration de San Francisco sur l'évaluation de la recherche : <https://sfdora.org/read/read-the-declaration-french/>

and get to know what is happening with their data » (comme mentionné à propos de la curation des données sur DaSCH (Gautschy, Subotic 2021)).

Mais plus généralement, on doit reconnaître et accorder aux communautés de recherche leur place centrale, et ce sont mêmes elles qui doivent donner l'impulsion pour développer et améliorer les pratiques en ORD :

« About stakeholder, there is a pressure coming from existing archive or service providers who are positioning themselves as drivers in OS and standardization process. But it is difficult because OS and the driver forces come from the research communities. We have to work on the communities and the international communities. The initial force always comes from the researchers. If we do not get the researchers on board, we will not get there. » (Ackermann, Schwede 2021)

Une fois initiée, la suite du processus s'appuie sur des personnes-clés:

« You identify people who are open-minded and willing to engage, from a good distribution of universities (including universities of applied sciences) and of disciplines. Try to make them reach consensus on metadata, data sharing, organizing training. » (Lutz 2021)

L'engagement des chercheurs dans la démarche se fait par entraînement et stimulation par les pairs:

« It is much more easier to get researchers enthusiastic by getting them in contact with other scientists than if told to do things because it is recommended by a librarian – the peer-to-peer effect (possibly someone from another field). » (Ackermann, Schwede 2021)

Agences de financement

Les agences de financement de la recherche ont aussi été citées plusieurs fois. Elles doivent soutenir financièrement les archives de données et les services nécessaires associés :

« What is really missing is the understanding by the funding agencies that once a community comes together with an agreement for standards and archive, they need to finance this archive [to support] the effort driven by the community for the high quality curation of the data, to make them admit that this kind of service adds value to the community. » (Ackermann, Schwede 2021)

mais aussi aller au-delà en soutenant la formation des chercheurs sur leurs compétences de gestion des données de la recherche:

« Funding agencies think they have solved the problem: there are now enough infrastructures and services. They are not aware of the whole process, what it really means. What is needed is to train the researchers and provide them with the skills needed to manage data. » (Gautschy, Subotic 2021)

Elles peuvent aussi être moteurs pour la mise en œuvre de grands principes (Déclaration DORA⁸) :

« Swissuniversities could work on implementing DORA in a better way and making ORD practices part of the evaluation criteria. » (Lutz 2021)

⁸ Déclaration de San Francisco sur l'évaluation de la recherche : <https://sfdora.org/read/read-the-declaration-french/>

En cohérence avec ce qui est discuté plus haut sur l'envergure de la recherche (qui si elle est européenne ou internationale est plus favorable aux communautés de recherche), on met également l'accent sur une coopération à l'échelle internationale des agences de financement pour assurer la durabilité des infrastructures d'archivage de données et le niveau de services associés :

« [Funding and sustainability of ORD repositories]: international research/national funding agencies: all funding agencies come together to create a global fund for Open data repositories. We need this different kind of mechanism to keep the archives alive. » (Ackermann, Schwede 2021).

« It is an international game with international communities. And one country (a national funding agency) will not be the curator of all structures. » (Ackermann, Schwede 2021).

Et sont invoqués plus généralement « all actors in the international context ». (Ackermann, Schwede 2021).

Bibliothécaires

On a discuté plus haut le rôle plus effacé des bibliothécaires par rapport à celui d'autres acteurs dans le soutien aux chercheurs pour effectuer certaines tâches caractéristiques de la GDR et de l'ORD.

T. Schwede et S. Ackermann leur attribuent bien un rôle, mais plus tourné vers la mise en place et l'organisation d'actions de sensibilisation et de formations pour les chercheurs. Le contenu serait assuré par des représentants de communautés très avancées en ORD :

« Education of the research communities about the advantages of OS and sharing...The library could organize the process [of training the researchers] ... [but not be] the one providing the content. Libraries often see the archives in a traditional way. It should be complemented by examples of best practices from other communities and fields that use different approaches than a library would do... Librarians may be asked to organize the process and to pull in researchers from communities to deliver the content. Librarians have lot of experience in the humanities field but when you talk about genomics or integrative modeling, you need people of the same field. » (Ackermann, Schwede 2021).

Infrastructures

Les représentants d'infrastructures interviewés (ou de structures associées, comme FORS pour FORSbase) les citent évidemment pour les services qu'ils fournissent aux chercheurs (en plus de l'archivage et de l'accessibilité des données). Ces services prennent la forme de fonctionnalités et de mécanismes standardisés (OLOS), ou de services disponibles pendant tout le cycle de vie des données (création, traitement, analyse des données par le personnel de FORS) :

« Develop services along all the data cycle for mandates or research projects who ask for support: creating, processing, analyzing data. A significant part of the funding and resources of FORS ... Staff of about 60, among them about 30-35 work on those services ... FORS is well positioned to provide services of interest for a lot of research institutions in Switzerland ... Our role is to make sure that the technical platform and services needed to support Open Research Data are in place and easy to use for the researchers to make the data available. » (Lutz 2021)

DaSCH quant à lui, contribue également à la sensibilisation et à la formation des chercheurs.

« DaSCH tries to cover some part of it [improvement of skills and better education at the individual scale] (with information, training, education...) » (Gautschy, Subotic 2021)

Institutions

Les politiques institutionnelles peuvent intervenir de façon complémentaire pour soutenir les communautés.

[réutilisation:] « Ce sont plutôt des pratiques qui doivent être développées dans les communautés de chercheurs. [Ce] sont plutôt les politiques institutionnelles qui peuvent agir et doivent mettre l'accent sur ce sujet. » (Burgi 2021)

Beate Böckem et Susanne Schumacher quant à elles, mentionnent le département de recherche de leur institution qui redirige les chercheurs vers la bonne personne (en l'occurrence ZHdK).

« The research affairs department helps with all that (the first contact point is in the department of BB). There is a contact person and he/she can help with questions, establish contact to the responsible person in the archive, in the library, or in the technical. » (Böckem et Schumacher 2021).

D'autres acteurs ont été cités : les développeurs, les éditeurs, les académies, le grand public.

Les développeurs : dans un exemple cité par T. Schwede et S. Ackermann, ils ont été mis à contribution pour développer un outil de génération automatique des métadonnées, permettant de limiter les erreurs faites manuellement sur les métadonnées et augmentant fortement la qualité de celles-ci.

« Discussed with the original software developers (only 10-15 software tools are used in the field) to agree on a mechanism with which the software automatically writes the metadata, without the user manipulating it. » (Ackermann, Schwede 2021).

« It also allowed to involve more the software community as an ally. » (Ackermann, Schwede 2021)

Les éditeurs: dans certains domaines, le matériel utile pour la reproduction des résultats de recherche est envoyé chez les éditeurs. Il faut établir et maintenir le dialogue entre dépôts et éditeurs de façon à ce que ce système ne menace pas le partage et la réutilisabilité des données.

« The FNS policy only asks for the replication material to be available, but FORS pushes for the original data and the documentation to be made available. The material which goes to publications is always a snapshot of it. FORS has set up a lighter service for the deposit of replication material: lighter metadata scheme, lighter documentation, shorter procedure, quality control. So that all replication material does not go to the publishing houses. » (Lutz 2021).

« Put services in place, keep the dialog with editors, push for reproducibility but not only in the publishers' system » (Lutz 2021)

Les académies: dans certains domaines de recherche, elles peuvent être un acteur important (coordination de projets de grande envergure, initiation de DaSCH, ...)

« Academies can play a role and do this in some extent in the medical field. The biggest project is the SPHN (Swiss Personalized Health Network) (besides SIB) that is administered by

the Academy of Medical Science. Academy of Social Sciences and humanities also played a role in humanities and in building DaSCH. » (Lutz 2021).

Le grand public: c'est l'« acteur final » du processus de développement d'une culture et de pratiques ORD, cité par T. Schwede et S. Ackermann. Mais son implication précise reste encore à expliciter :

« An open data or open science repository with a certain value also becomes relevant for the general public and for education. Open science data = valuable to the society as a whole. We should not forget those stakeholders in that context. » (Ackermann, Schwede 2021)

B. Stratégie d'implémentation

La création et le développement de communautés sont donc reconnus nécessaires (éventuellement au prix de certains efforts):

« You need to invest a lot in community buildings. You have to understand all these projects as community-building projects and that is the hard part. » (Lutz 2021)

A cette fin deux types d'approches sont évoqués lors des entretiens.

Une approche bottom-up

En ce qui concerne la création et le développement d'une communauté, une approche bottom-up est préconisée. Elle est mentionnée pour différents aspects :

Enclenchement du processus

Il est reconnu que l'impulsion doit forcément venir des chercheurs eux-mêmes :

« OS and the driver forces come from the research communities. We have to work on the communities and the international communities. The initial force always comes from the researchers. If we do not get the researchers on board, we will not get there. » (Ackermann, Schwede 2021).

Fonctionnement

La gouvernance doit se mettre en place également au niveau de la communauté (ou sous-communauté) :

« The community-building needs to be steered and governed within each of the sub-community » (Lutz 2021)

La définition des standards doit aussi venir des communautés ou éventuellement se faire en prenant en compte l'ensemble des besoins de la communauté :

« Standards [et c'est aux communautés de les amener]. » (Burgi 2021)

« Metadata schemes, standardization of formats, ... Standardization is crucial. Need of at least one better data scheme per discipline, acceptance of how things should be documented and described, in a disciplinary way. It is where the community-building aspect comes into place. Not wise to impose standards. Create a frame, a structure where the dialog happens. » (Lutz 2021).

« [Data harvesting]: it was really worthwhile bringing the community together and agreeing on standards, going to the source of the software and not relying on the individual scientists. » (Ackermann, Schwede 2021)

[Referring to Helen Berman's paper on building the PDB community]: « To agree on common standards on data archiving and sharing. The PDB made sure to pick up all the needs of the communities to further develop the archive. » (Ackermann, Schwede 2021)

Par discipline

Une approche disciplinaire peut être mentionnée pour certains aspects :

« The community-building and the service development must happen in a disciplinary way. But the technical infrastructure does need to be disciplinary focused. Swissuniversities does not have this disciplinary focus, FNS has it much more, but swissuniversities could formulate the needs of the universities. » (Lutz 2021)

« Services that need to be discipline-specific (metadata schemes), but not necessarily tools and technical infrastructure » (Lutz 2021)

« It is also good for the disciplines to have a gentleman agreement (or other) in the community saying that nobody publishes before the first study author (the one who generated data) has published. » (Ackermann, Schwede 2021)

Une approche top-down

Elle est aussi préconisée pour favoriser le fonctionnement et le développement des communautés, au niveau de la stratégie :

« Aux USA, Obama avait imposé que tout soit en accès libre au bout de tant d'années et maintenant les grandes infrastructures de recherche l'appliquent. Mais il ne faut pas trop compter sur l'engagement au niveau des chercheurs. Il faut toujours penser cela en top-down. » (Burgi 2021)

Et pour l'application des principes FAIR :

« The big topic is not so much as developing tools as connecting tools. We need to make the whole system of data infrastructure FAIR. » (Lutz 2021)

Plusieurs citations évoquent le besoin d'établir et de maintenir la connexion de différents acteurs entre eux, en-dehors du cadre strict de la communauté de recherche, avec éventuellement des difficultés :

« Put services in place, keep the dialog with editors, push for reproducibility but not only in the publishers' system. » (Lutz 2021)

« It might be difficult for swissuniversities to play a coordinating role if the funding mechanism is completely disconnected from the coordination role. » (Lutz 2021)

« Swissuniversities could work on implementing DORA in a better way and making ORD practices part of the evaluation criteria. Mandate from the SERI to develop a strategy in the national plan, and to have coordination. But coordination aspects are not clear. FNS put up a

reflection group on data infrastructure, but without swissuniversities or representatives of academies. » (Lutz 2021).

Deux citations font ressortir des limitations à cette approche top-down:

« Constant lack of coordination in Switzerland overall, very little top-down coordination, a big issue » (Lutz 2021)

« Top-down model is quite expensive and hardly extendable at a large scale. » (Clivaz 2021)

C. Sensibilisation, éducation et formation

Après l'identification des acteurs-clés et le choix de la stratégie la plus pertinente pour la mise en place et le développement de communautés de recherche, un troisième aspect pour faciliter et développer la culture et les pratiques de l'ORD est apparu au cours des entretiens :

« Educate people - education - education – education » (Gautschy, Subotic 2021)

Cette « education » englobe à la fois la sensibilisation aux bénéfices de l'ORD :

« Education also includes making the researchers fully aware that sharing and re-using data (the possibilities of the semantic web, ...) is in his/her own interest. » (Gautschy, Subotic 2021)

« Educate more the research communities about the advantages of OS and sharing and what it means and how it could work. Librarians may be asked to organize the process and to pull in researchers from communities to deliver the content (peer-to-peer effect). » (Ackermann, Schwede 2021)

« Training workshops will allow large discussion in Humanities and convince scholars to produce data with a good balance between flexibility and standardization » (Clivaz 2021)

et la formation adressée aux chercheurs pour l'amélioration des compétences à maîtriser pour de bonnes pratiques ORD :

« People need more skills and a better education at the individual scale. The education part is the most important. You can always offer more, but without education, providing more services would be useless. » (Gautschy, Subotic 2021)

« What is needed is train the researchers and provide them with the skills needed to manage data. » (Gautschy, Subotic 2021)

« Cela passe par la formation au niveau suisse sur la GDR » (Burgi 2021)

« SNSF/ORD joint workshop would be wishable. » (Clivaz 2021)

Des précisions sont données sur la temporalité et le public-cible de ces actions de sensibilisation et de formation.

Il faut intervenir le plus vite possible dans le processus de recherche, avant l'étape de collecte de données :

« You can optimize by teaching the researchers before they start collecting data » (Lutz 2021)

« Training of researchers: Try to capture the people as early as possible [as soon as the project is funded]. It should be a joint effort between research funders and universities to put support in place and help the researchers. » (Lutz 2021)

mais aussi former les futurs chercheurs alors qu'ils sont encore étudiants:

« Training of students » (Lutz 2021)

« Universities can educate students because they need it in their future » (Gautschy, Subotic 2021)

« Whatever the disciplines, at the Master level (or even at the Bachelor level), there should be mandatory basic training. The students there are the future application writers. » (Gautschy, Subotic 2021)

« About older researchers: some of them are interested; for other ones, it is hopeless. They tend to delegate the management tasks to their students. » (Gautschy, Subotic 2021)

IV. Recommandations pour développer le niveau de pratiques de l'ORD au sein des communautés

Les recommandations présentées dans cette partie sont issues des constats faits durant les entretiens et des résultats de l'enquête. Dans le but de rapprocher au mieux la stratégie des pratiques, elles suivent une logique de bottom-up et de top-down comme conseillé.

1) « Talking the same language and starting from the beginning »

// Définir le concept de « communauté de recherche »

Le terme "communauté de recherche" est le concept central de ce travail, comme demandé par le mandataire; il est également mentionné dans la revue de la littérature. Les résultats de ce travail confirment le potentiel de ces communautés dans le développement des compétences ORD des chercheurs. Cependant, nos résultats soulignent également qu'il y a encore peu de compréhension et de consensus sur ce terme parmi les experts OS/ORD et parmi les chercheurs. Les données recueillies lors des entretiens et de l'enquête montrent que le terme "communauté de recherche" est souvent compris comme la communauté de la discipline de recherche elle-même, même si nous avons proposé d'autres définitions lors de la collecte des données. Le terme "communauté de données" discuté dans la littérature est encore moins utilisé par les chercheurs, comme le soulignent les résultats de l'enquête massive.

Avant de chercher à renforcer la culture de l'ORD et à améliorer les pratiques de l'ORD des chercheurs, il faudrait donc s'efforcer de comprendre ce qu'est une communauté de recherche ou de données et comment elle aide les chercheurs à améliorer leurs pratiques ORD ; et à partir de là, construire et consolider ces communautés.

Recommandations :

- **1a** S'accorder sur une définition des communautés de recherche qui inclut leur variété mais précise l'importance d'une organisation formelle (ex : par projets, sous forme d'associations, autour d'infrastructures...)
- **1b** S'appuyer sur des modèles d'organisation formelle qui permettent d'identifier des points de contacts (personnes en charge, institutions...)

- **1c** Répertorier les communautés et les points de contact qui permettent d'échanger avec elles

2) Accompagner la construction et la consolidation des communautés de recherche

Initier la formation d'une communauté, la consolider et la soutenir dans l'acquisition de la maturité en matière d'ORD est un processus à double sens. Nos résultats soulignent qu'il ne sert à rien de la forcer si aucune motivation forte ne vient des chercheurs eux-mêmes. Cependant, elle nécessite un cadre de gouvernance stable, en connexion avec un environnement plus large, adapté aux différents contextes, ainsi qu'une planification à long terme.

Recommandations :

- **2a** Reconnaître la nécessité d'une approche diamétrale (bottom up et top down)
- **2b** Mettre en place des outils qui permettent aux chercheurs de faire remonter leurs besoins ou leurs questions et de les impliquer dans la suite du processus

3) Développer les mesures d'incitation (amoindrir les obstacles)

- **3a** Reconnaître à leur juste valeur l'effort, le temps et les compétences que les chercheurs investissent dans de bonnes pratiques ORD : prendre en compte ces aspects pour le recrutement dans l'institution et l'avancement de la carrière académique
- **3b** Favoriser l'usage d'accords informels par communauté pour une éthique de partage des données : pas de publication d'article avant que le premier auteur de l'étude ou le premier producteur de données n'ait lui-même publié

4) Impliquer les différents acteurs et formaliser leur contribution

Plusieurs acteurs jouent des rôles différents dans ce contexte et devraient être impliqués à des étapes spécifiques du processus. La section III présente ces acteurs potentiels et leur implication possible à partir des résultats des entretiens et de l'enquête. Ce qui ressort également de ce mandat est le fort intérêt de nombreuses personnes pour ce sujet et une envie de partager leurs connaissances.

Recommandations

- **4a** Identifier des personnes particulièrement actives sur le sujet (spécialistes ou membres de communautés avancés dans leur pratique et dans la culture ORD) susceptibles de fédérer d'autres membres
- **4b** Formaliser un réseau de ces personnes entre elles (transposer le modèle de Data Champions à « Community Champions »)

5) Utiliser la communication et la formation en temps opportun

La communication est essentielle pour sensibiliser aux avantages de la pratique ORD. En ce sens, elle devrait être prioritaire lorsqu'on essaie de soutenir la formation d'une nouvelle communauté.

La formation est un élément fondamental du processus. Elle ne répond pas aux mêmes objectifs selon l'étape du processus mais doit en faire partie tout au long, et, au-delà de la consolidation de la communauté, pour qu'elle puisse exister à long terme.

Cependant, le manque de temps est une réelle problématique qui apparaît dans les réponses des chercheurs. Proposer des journées entières de formation est contradictoire quand le manque de temps est déjà un obstacle en soi.

Recommandations

- **5a** Proposer des formats de formation de format très court (micro-learning) et facilement accessibles
- **5b** Former les acteurs de l'ORD à cette méthodologie de formation
- **5c** Faire intervenir les représentants des Community Champions

6) Fournir le bon niveau de soutien pour la définition et l'adoption des normes

La standardisation est un aspect essentiel de la maturité de l'ORD. L'existence et l'utilisation de normes par les communautés est un indicateur de la maturité de leurs pratiques de ORD. L'adoption de normes peut également accélérer la motivation des chercheurs à suivre de bonnes pratiques de ORD s'ils en voient les avantages. Cependant, il ne sert à rien d'imposer ces normes. Elles doivent être conçues et mises en œuvre par les membres de la communauté eux-mêmes, sinon il y a de fortes chances qu'elles ne soient pas pleinement adoptées par la communauté. Le soutien doit se faire d'une autre manière : identifier les membres de la communauté susceptibles de fédérer d'autres membres autour de la définition des standards et leur apporter un soutien adéquat tout au long du processus.

Recommandations :

- **6a** Encourager les projets qui visent la standardisation avec une approche bottom-up
- **6b** Communiquer et valoriser l'utilisation de standards utilisés et/ou développés

7) Planifier un financement durable

Le manque de temps a été largement évoqué durant ce travail et au final sa prise en compte doit se faire, notamment au point de vue financier. Ce financement doit être suffisamment durable pour permettre aux infrastructures, aux services et aux activités de sensibilisation et d'accompagnement (communication, formation et soutien) d'être disponibles tout au long du processus de développement de la maturité de la communauté en ORD, et au-delà, pour pérenniser la communauté et ses bonnes pratiques. Lors du soutien des premiers pas d'une nouvelle communauté, le financement devrait être fourni au cas par cas, pour répondre aux besoins spécifiques qui se présentent. Dans les étapes suivantes, il devrait être fourni régulièrement pour garantir la durabilité à long terme des infrastructures et des services mis à la disposition des communautés.

Recommandations :

- **7a** Evaluer les ressources financières et humaines nécessaires à chaque étape du processus de recherche (pourcentage à établir en lien avec la typologie de la recherche)
- **7b** Prendre en compte cette évaluation dans l'estimation des financements nécessaires des projets de recherche

Matrice de décision

Nous avons attribué un numéro d'identifiant à chaque recommandation (16 mesures au total) et les avons regroupées dans un Tableau 7.

Tableau 7 : Identifiants des recommandations pour la matrice de décisions

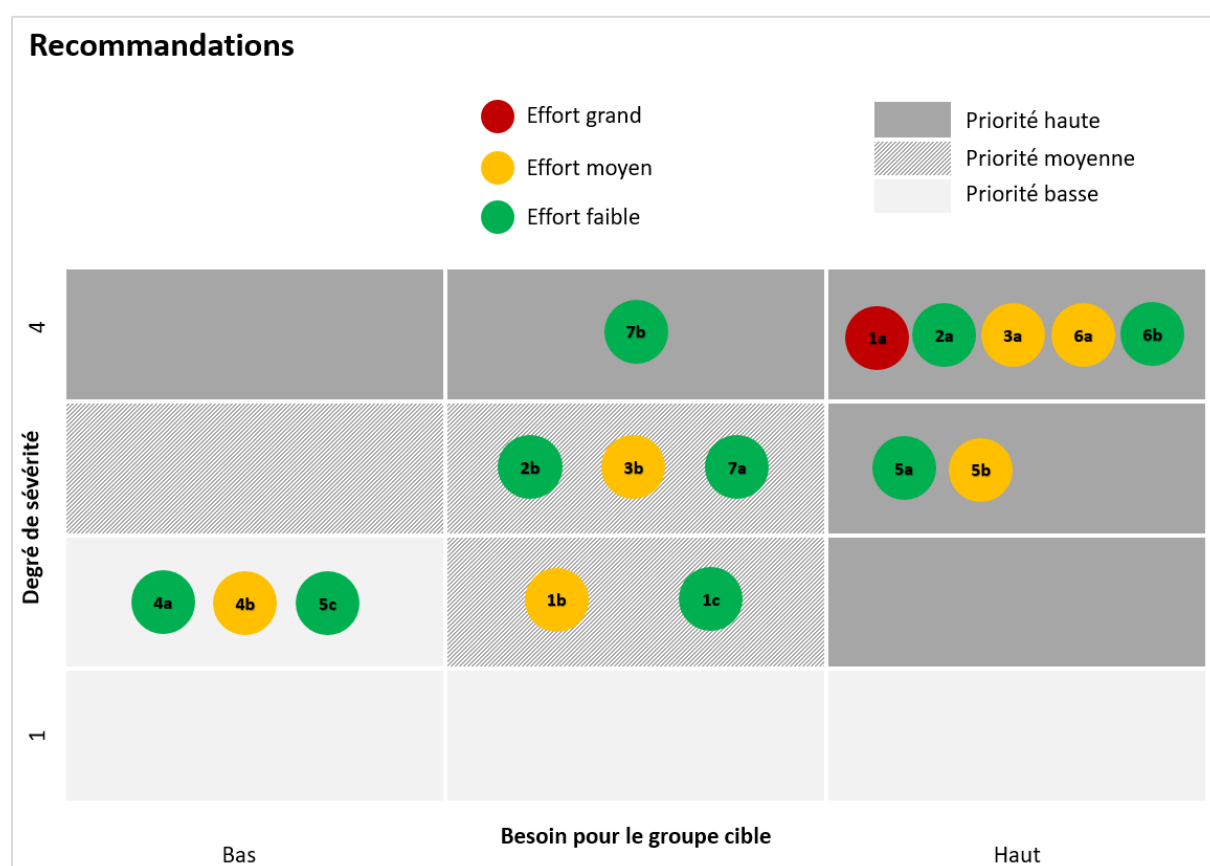
Identifiant	Recommandations
1a	S'accorder sur une définition des communautés de recherche qui inclut leur variété mais précise l'importance d'une organisation formelle (ex : par projets, sous forme d'associations, autour d'infrastructures...)
1b	S'appuyer sur des modèles d'organisation formelle qui permettent d'identifier des points de contacts (personnes en charge, institutions...)
1c	Répertorier les communautés et les points de contact qui permettent d'échanger avec elles
2a	Reconnaître la nécessité d'une approche complémentaire (bottom up et top down)
2b	Mettre en place des outils qui permettent aux chercheurs de faire remonter leurs besoins ou leurs questions et de les impliquer dans la suite du processus
3a	Reconnaître à leur juste valeur l'effort, le temps et les compétences que les chercheurs investissent dans de bonnes pratiques ORD : prendre en compte ces aspects pour l'engagement dans l'institution et l'avancement de la carrière académique
3b	Favoriser l'usage d'accords informels par communauté pour une éthique de partage des données : pas de publication d'article avant que le premier auteur de l'étude ou le premier producteur de données n'ait lui-même publié
4a	Identifier des personnes particulièrement actives sur le sujet (spécialistes ou membres de communautés avancés dans leur pratique et dans la culture ORD) susceptibles de fédérer d'autres membres
4b	Formaliser un réseau entre eux (transposer le modèle de Data Champions à « Community Champions »)
5a	Proposer des formats de formation de format très court (micro-learning) et facilement accessibles
5b	Former les acteurs de l'ORD à cette méthodologie de formation
5c	Faire intervenir les représentants des Community Champions
6a	Encourager les projets qui visent la standardisation avec une approche bottom-up
6b	Communiquer et valoriser l'utilisation de standards utilisés et/ou développés

7a	Evaluer les ressources financières et humaines nécessaires à chaque étape du processus de recherche (pourcentage à établir en lien avec la typologie de la recherche)
7b	Prendre en compte cette évaluation dans l'estimation des financements nécessaires des projets de recherche

La matrice de décision permet de visualiser le niveau de priorité pour chaque recommandation, selon son degré de sévérité et le besoin du groupe cible (Figure 32). On évalue aussi le niveau d'effort nécessaire pour les implémenter :

- Priorité : basse, moyenne et haute
- Sévérité de recommandation: le niveau d'importance de la recommandation (qui influe donc sur l'impact de cette dernière). 1 étant le degré de sévérité le plus bas et 4 le plus élevé.
- Besoin pour le groupe cible: bas, moyen et haut
- Effort : faible, moyen et grand

Figure 32 : Matrice de décision



Conclusion

Swissuniversities, dans le cadre de sa Stratégie nationale pour des données de recherche ouvertes initiée par sa Délégation Open Science, a mandaté la HEG et la FHGR pour réaliser deux études complémentaires. La première doit caractériser le paysage des données de recherche ouvertes (ORD) en Suisse et la deuxième, réalisée par la HEG, doit caractériser les pratiques et les besoins des chercheurs en ORD en se focalisant sur les communautés de recherche.

Pour réaliser ce travail, deux phases se sont succédées. Tout d'abord, ont été organisés des entretiens semi-structurés de représentants d'infrastructures de partage de données et de communautés de recherche. Ensuite, à l'aide d'un questionnaire en ligne, les chercheurs académiques suisses ont été interrogés. Ces deux étapes visaient à collecter des informations sur la culture et les pratiques actuelles en ORD, ainsi que les besoins d'amélioration pour ces pratiques, et les acteurs à impliquer.

Le rapport se déploie en quatre grands axes, chacun déroulant ses propres parties.

Le premier axe s'intéresse plus globalement aux « communautés de recherche » et discute les différentes définitions qui sont faites de ces dernières, leur organisation ainsi que deux profils particuliers identifiés, les membres des communautés de données et les chercheurs « sans données ».

Le deuxième axe cible le rôle de l'organisation en « communautés de recherche » pour les pratiques ORD et révèle les obstacles ainsi que les facteurs de motivations pour le développement ORD. Il éclaire sur l'utilisation de standards ainsi que sur le niveau de compétences en ORD. Dans la perspective d'une élaboration future d'un modèle de niveau de maturité, ces deux aspects seraient des éléments essentiels à prendre en compte.

Le troisième axe met en évidence les leviers possibles pour contribuer au développement et à l'avancement des pratiques ORD. Il s'agit notamment d'impulser la création de communautés de recherche en s'appuyant sur des approches simultanées bottom up et top down, des actions de sensibilisation, d'éducation et de formation, ainsi que sur les acteurs-clés essentiels. Il ressort que la communauté, l'institution, et l'infrastructure de données apportent un soutien notable aux chercheurs pour de bonnes pratiques en ORD. Les bibliothécaires pourraient être impliqués davantage dans des actions de sensibilisation et pour l'organisation de formations dédiées.

Le quatrième axe, enfin, propose des recommandations pour le développement du niveau de pratiques ORD au sein des communautés. Elles se déclinent en sept parties qui mettent chacune en avant des recommandations spécifiques. Une matrice de décision permet de planifier leur mise en place stratégique.

Les conclusions de ce mandat valident l'hypothèse que l'appartenance à une communauté de recherche a un impact sur le niveau de développement des pratiques ORD des chercheurs. Elles permettent également d'identifier des concepts et des éléments utiles en vue de l'élaboration d'un modèle détaillé pour l'évaluation du niveau de maturité en ORD des communautés de recherche. Ce modèle serait un outil utile pour le suivi et le soutien au développement de l'ORD et de l'OS de la recherche académique en Suisse.

Bibliographie

BERMAN, Helen M., KLEYWEGT, Gerard J, NAKAMURA, Haruki et MERKLEY John L, 2013. How Community Has Shaped the Protein Data Bank. *Structure* [en ligne]. 3 septembre 2013. 21(9), pp. 1485-1491. [Consulté le 11.06.2021]. Disponible à l'adresse : <https://doi.org/10.1016/j.str.2013.07.010> [accès par abonnement]

BÖCKEM, Beate, SCHUMACHER, Susanne, 2021. Entretien avec Mme Böckem, Head of research affairs at ZHdK, et Mme Schumacher, Co-chair of the digital council of ZHdK, visioconférence, 2 juillet 2021.

BURGY, Jean-Yves, 2021. Entretien avec Pierre-Yves Burgy, Directeur adjoint de la division des systèmes d'information de l'UNIGE, initiateur du projet OLOS et co-directeur du DLCM, visioconférence, 21 mai 2021.

COOPER, Danielle, SPRINGER, Rebecca, 2019. Communautés de données : Communautés de données: A New Model for Supporting STEM Data Sharing. Ithaka S+R [en ligne]. 13 Mai 2019. Disponible à l'adresse <https://doi.org/10.18665/sr.311396>

COOPER, Danielle, SPRINGER, Rebecca, 2020. Communautés de données: Empowering Researcher-Driven Data Sharing in the Sciences. *International Journal of Digital Curation* [en ligne]. 30 décembre 2020. Vol. 15, No 1. [Consulté le 11.06.2021]. Disponible à l'adresse : <https://doi.org/10.2218/ijdc.v15i1.695>

LUTZ, Georg, 2021. Entretien avec Georg Lutz, directeur de FORS, visioconférence, 17 juin 2021.

Research Data Alliance FAIR Data Maturity Model Working Group, 2020. FAIR Data Maturity Model: Specification and Guidelines. 1er juin 2020. Disponible à l'adresse: <https://doi.org/10.15497/rda00050>

SCHWEDE, Torsten, 2021. Entretien avec M. Torsten Schwede, vice-président de la recherche, chef du groupe de recherche Computational Structural Biology Group, Université de Bâle, Institut suisse de bioinformatique, visioconférence, 28 mai 2021.

Swissuniversities, 2021. Stratégie Nationale Suisse Open Research Data. Juillet 2021. Disponible à l'adresse: <https://www.swissuniversities.ch/fr/themes/digitalisation/open-research-data>

V. Annexes

Annexe 1: Concepts-clés

Les concepts utilisés dans le cadre de ce travail sont définis ainsi:

Communautés de données (Data Communities): selon Danielle Cooper et Rebecca Springer: "[Formal] or informal groups of scholars who share a certain type of data with each other, regardless of disciplinary boundaries" (Cooper, Springer, 2019).

Communautés de recherche : réseaux de chercheurs partageant le même (type de) données, les mêmes outils et infrastructures de partage de données, les mêmes pratiques de RDO et la même culture de l'Open Science, mais n'appartenant pas nécessairement à la même discipline de recherche.

Données de recherche : selon la stratégie nationale suisse des données de recherche ouvertes : "Les données de recherche sont des éléments factuels collectés, observés ou générés qui sont communément acceptés dans la communauté scientifique comme nécessaires pour permettre, documenter ou valider les résultats de la recherche." (swissuniversities 2020).

Enquête (en ligne) : méthode de collecte d'informations par écrit. Elle permet de cibler un grand nombre de destinataires et de recevoir leurs réponses dans un laps de temps limité.

- Entretien (semi-structuré) : méthode de collecte verbale d'informations. Dans les entretiens semi-structurés, la liste des thèmes à aborder et certaines questions à poser sont fixées. Les autres questions et l'ordre dans lequel elles sont posées restent flexibles, ce qui permet de s'adapter à la personne interrogée (pour autant que la discussion suive un ordre naturel).

Infrastructures : les infrastructures de partage de données, à savoir les solutions de stockage et de partage des données. Au-delà du minimum, elles fournissent très souvent des fonctionnalités et des outils spécifiques pour aider le processus de dépôt, rendre les données FAIR et les archiver. Par défaut, le terme "infrastructures" inclura également ces fonctionnalités et outils.

Services de soutien: services fournis aux chercheurs pour adopter et améliorer leurs pratiques de gestion des données de recherche. Ces services peuvent inclure des informations, des formations, du coaching, etc.

Annexe 2: Modèle de guide pour les entretiens semi-structurés

Scope of the interview

Introduction and short reminder of the objectives – 5 mn

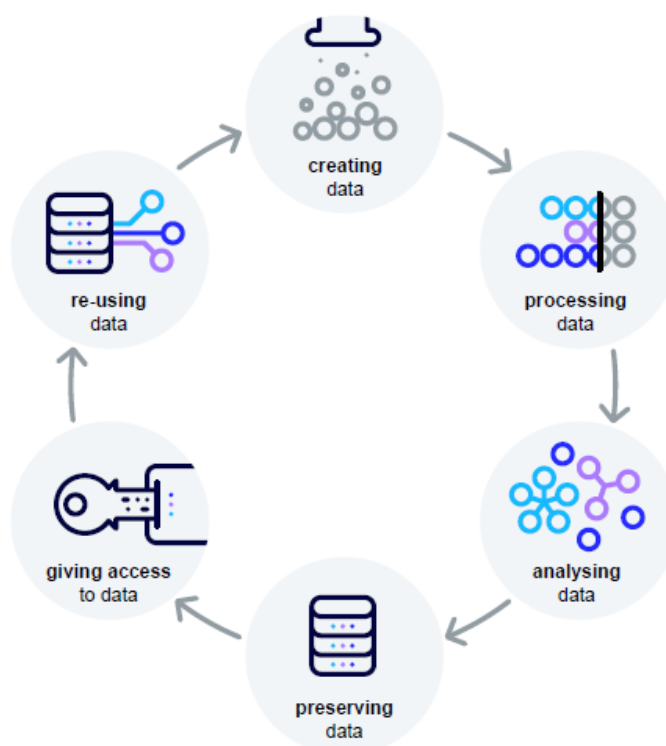
- I. Infrastructures and support services for ORD – open questions – 15 mn
- II. Maturity level of ORD practices - card sorting – 15 mn
- III. Identified needs for ORD – open questions – 15 mn

Conclusion (validation procedure of the interview report and reminder about the survey to come) – 5 mn

Detailed content

Infrastructures and support services for ORD - open questions

1. What kind of infrastructure (data sharing repository) and support services does your research group use for managing and sharing its data ? Why ?
2. In the Data Lifecycle Management Cycle illustrated as below, which steps is more relevant to your research group, which steps does it practice more?



3. About « re-using data » (according to the answer given at the previous question) :

Option 1 – if the « re-using » step is stated (besides giving access) : in which way do you think your research group facilitates research reproducibility (besides sharing data) ?

Option 2 - if the « re-using » step is not stated : are there plans to contribute more to research reproducibility, and if yes, in which way ?

4. Within research in bioinformatics or at your institution, do you identify specific research data communities ? If yes, which ones ? Are there researchers in those communities that we could contact on your behalf to investigate more about their practices and their needs (in terms of ORD)?

Note : By research communities, we mean networks of researchers sharing the same (type of) data, the same data sharing tools and infrastructures, the same ORD practices, and the same culture of Open Science, but not necessarily being from the same research discipline. We are interested in getting your views on this definition, if you agree with it or if you have suggestions for another one.

Maturity level of ORD practices – card sorting

Guidelines :

- Sort out the communities according to how you evaluate their ORD level: low, intermediate or high.
- On which criteria do you base for sorting the cards?

Identified needs for ORD – open questions

1. About research data, can you think of other types of data than the following ones:

- observations
- experiences
- simulations
- derived data
- references
- digitalizations (see details in Appendix 1) ?

2. If yes, which types ? In which disciplines or research communities ? Does the management of these data follow different steps from the ones: “preserve, give access, re-use” ?

Note : we are trying here to identify research communities with less common views or behaviour in terms of data management. Better understanding their practices may allow to address their needs with more adapted support (infrastructures, services, ...).

3. In a general way of speaking, according to you, what are the main obstacles for ORD (and more specifically for data re-using) ? Are they different from one research community to the other ?

4. What are the main stimulating factors for ORD (and data re-using) ? Are they different from one research community to the other ?

5. In view of the current offer of infrastructures, services, incentives available to the researchers, what are their needs to more actively engage in ORD and to improve their practices (and more specifically relating to data re-using)? Do those needs differ by research communauté de données ?

6. Do you have other suggestions to overcome the current shortcomings or caveats and to further contribute to better practices and more commitment into ORD?

7. In a general way, which shareholders could be more involved ?

Bibliography

RITZE, Dominique, ECKERT, Kai and PFEFFER, Magnus, 2013. *Forschungsdaten*.
In: DANOWSKI, Patrick, (Open) Linked Data in Bibliotheken. Berlin : De Gruyter Saur, 2013, pp. 122-138. ISBN 978-3-11-027634-3

Typology of Research Data

Art	Characteristik	Beispiel
Beobachtungen	Daten werden in Echtzeit erfasst meistens unersetzbar	Sensordaten Umfragedaten
Experimente	meist im Labor erstellt reproduzierbar aber teuer	Gensequenzen Chromatogramme
Simulationen	von Testmodellen generiert Model und die Metadaten wichtiger als Ausgabe	Klimamodelle Wirtschaftsmodelle
Abgeleitete Daten	aus anderen Daten abgeleitet oder kompiliert reproduzierbar	Textmining 3D-Modelle
Referenzen	Sammlung kleinerer Datensätze Meist publiziert	Gensequenzdatenbank Primäre Textquellen
Digitalisierungen	Digitale Versionen von analogen Objekten, reproduzierbar, solange die Originale verfügbar sind	Manuskripten

(Adapted from Ritze et al. 2013)

Annexe 3 : Thèmes et sous-thèmes de la grille d'analyse des données collectées avec les entretiens

Topic	Sub-topic
Disciplines and communities	
	Individual/Communities
	Difference between disciplines
	Difference between disciplines - reasons
Organization of communities	
	Culture of reusing - citation
	Culture of reusing - topic
Typology of research and consequences on ORD	
	Types of data
	Methodology
	Research time scale
	Research level (national, international, ...)
	Publication
	ORD – bottom up, building communities
	ORD – top down
Concerns – obstacles - incentives	
	Concerns - citation
	Concerns - topic
	Obstacles - citation
	Obstacles - topic
	Incentives - citation
	Incentives - topic
Solutions for ORD development	
	Standardization - citations
	Standardization - topics
	Education
	Skills needed
About infrastructure	
	Motivation and goal
	Cycle
	Functionality
	Standard
	Statistic
Other topics	
	Licenses
	Metadata
	“False” obstacles
	Actors
	Public open repository – commercial structures

Annexe 4: Contenu détaillé de l'enquête en ligne

Enquête OpenScience

Introduction

This survey is performed within the framework of a swissuniversities mandate and in the context of the National Open Research Data Strategy initiated by the Open Science Delegation of swissuniversities. The mandate specifically aims at characterizing the practices and needs in ORD within Swiss research communities and **the survey has been designed for researchers** based in Switzerland.

If you are not a researcher yourself but in a position where you have access to a network of researchers in Switzerland (e.g. OA/ORD/OS officer or librarian at Swiss higher education institutions), please forward it via this network.

This survey will stay open until 29 August 2021. **It should take about 15-20 minutes to complete.** The data collected will be used only for this mandate. We thank you in advance for your contribution to our work.

HEG mandate team

Dir : Prof. René Schneider

Haute Ecole de Gestion de Genève
Filière Information documentaire
Rue de la Tambourine 17 - 1227 Carouge
gaia.bongi(at)hesge.ch
marielle.guirlet(at)hesge.ch
elise.pelletier(at)hesge.ch

Fin de bloc: Introduction

Début de bloc: Question pour tous

A1 What is your institution of affiliation?

- Federal institute (1)
- University (2)
- University of Applied Sciences (3)
- University of Teacher Education (4)
- Higher Education Institution (5)
- Accredited institution (6)

Afficher cette question :

If A1 = Federal institute

A2 - 1 Please specify your answer given to the last question:

- Ecole polytechnique fédérale de Lausanne EPFL (1)
- Eidgenössische Technische Hochschule Zürich ETHZ (2)
- Swiss Federal Institute for Vocational Education and Training, SFIVET (3)
- Swiss Federal Institute of Sport Magglingen, SFISM (4)
- Paul Scherrer Institute, PSI (5)
- Swiss Federal Institute of Aquatic Science and Technology, Eawag (6)
- Swiss Federal Laboratories for Materials Science & Technology, Empa (7)
- Swiss Federal Institute for Forest, Snow and Landscape Research, WSL (8)

Afficher cette question :

If A1 = University

A2 - 2 Please specify your answer given to the last question:

- Universität Basel (1)
- Universität Bern (2)
- Université de Fribourg (3)
- Université de Genève (4)
- Université de Lausanne (5)
- Universität Luzern (6)
- Université de Neuchâtel (7)
- Universität St. Gallen (8)
- Università della Svizzera italiana (9)
- Universität Zürich (10)

Afficher cette question :

If A1 = University of Applied Sciences

A2 - 3 Please specify your answer given to the last question:

- BFH - Bern University of Applied Sciences (1)
- FHGR - University of Applied Sciences of the Grisons (8)
- FHNW - University of Applied Sciences and Arts Northwestern Switzerland (2)
- HES-SO - University of Applied Sciences and Arts Western Switzerland (3)
- HSLU - Lucerne University of Applied Sciences and Arts (5)
- Kalaidos FH - University of Applied Sciences (4)
- OST - Eastern Switzerland University of Applied Sciences (6)
- SUPSI - University of Applied Sciences and Arts of Southern Switzerland (7)
- ZFH - Zurich University of Applied Sciences and Arts (9)
- ZHAW - Zurich University of Applied Sciences (11)
- ZHdK - Zurich University of the Arts (12)
- PHZH - Zurich University of Teacher Education (13)
- HWZ - Zurich University of Applied Sciences in Business Administration (14)

Afficher cette question :

If A1 = University of Teacher Education

A2 - 4 Please specify your answer given to the last question:

- Haute école pédagogique des cantons de Berne, du Jura et de Neuchâtel (1)
- Haute école pédagogique du canton de Vaud (2)
- Pädagogische Hochschule Wallis | Haute école pédagogique du Valais (3)
- Haute école pédagogique Fribourg | Pädagogische Hochschule Freiburg (4)
- Interkantonale Hochschule für Heilpädagogik Zürich (5)
- Pädagogische Hochschule Graubünden | Alta scuola pedagogica dei Grigioni | Scola auta da pedagogia dal Grischun (6)
- Pädagogische Hochschule Bern (7)
- Pädagogische Hochschule Luzern (8)
- Pädagogische Hochschule Nordwestschweiz (FHNW) (9)
- Pädagogische Hochschule St. Gallen (10)
- Pädagogische Hochschule Schaffhausen (11)
- Pädagogische Hochschule Schwyz (12)
- Pädagogische Hochschule Thurgau (13)
- Pädagogische Hochschule Zürich (14)
- Pädagogische Hochschule Zug (15)
- Schweizer Hochschule für Logopädie Rorschach SHLR (16)
- SUPSI - Dipartimento formazione e apprendimento (Ticino, SUPSI) (17)

Afficher cette question :

If A1 = Higher Education Institution

A2 - 5 Please specify your answer given to the last question:

- Graduate Institute of International and Development Studies, Genève, IHEID (1)
- Stiftung Universitäre Fernstudien Schweiz, Brig / Fondation Formation universitaire à distance, Suisse à Brigue (2)

Afficher cette question :

If A1 = Accredited institution

A2 - 6 Please specify your answer given to the last question:

- Facoltà di Teologia di Lugano (1)
- Franklin University Switzerland (2)
- Staatsunabhängige Theologische Hochschule Basel (3)
- Theologische Hochschule Chur (THC) (4)

A3 To which research field(s) or discipline(s) does your research relate?

- Theology & religious studies, history, classical studies, archaeology, prehistory and early history (1)
- Linguistics and literature, philosophy (28)
- Art studies, musicology, theatre and film studies, architecture (29)
- Ethnology, social and human geography (26)
- Psychology, educational studies (27)
- Sociology, social work, political sciences, media and communication studies, health (24)
- Economics, law (25)
- Mathematics (22)
- Astronomy, Astrophysics and Space Science (23)
- Chemistry (30)
- Physics (2)
- Engineering Sciences (3)
- Environmental Sciences (4)
- Earth Sciences (5)
- Basic Biological Research (6)
- General Biology (7)
- Basic Medical Sciences (8)
- Experimental Medicine (9)
- Clinical Medicine (10)
- Preventive Medicine (Epidemiology/EarlyDiagnosis/Prevention) (11)
- Social Medicine (32)
- Other (33)

A4 Please specify your answer given to the last question:

A7 Would you say that you are part of a **research community** (or of several ones)?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If A7 = Yes

A8 Which one(s)?

Afficher cette question :

If A7 = Yes

A9 How would you describe this "research community"?

- Community related to data (data research community) (1)
- Community related to a project (2)
- Community related to a research lab (3)
- Community related to a specific research discipline or field (4)
- Community related to an infrastructure (5)
- Other: (6) _____

Fin de bloc: Question pour tous

Début de bloc: With (data) communities (A5 or A7 = YES)

Characterization of community and research

B1 How is your community organised?/How are your communities organised?

- informally (1)
- in connection with an infrastructure (2)
- in connection with a project (3)
- as an association (4)
- on a social network (5)
- Other: (6) _____

B2 Is there a website for one or more of your communities?

- No (2)
- Yes (1)

Afficher cette question :

If B2 = Yes

B3 Please give the website address. For the rest of the survey, please focus on one community in which you are the most active (enter the website of this community in first position).

B4 What is the size of your community (approximate number of researchers)?

B5 What is the typical scale of research in your community?

- National (1)
- European (2)
- International (3)
- Other (4) _____

B6 What is the typical time frame of research in your community?

- less than 1 year (1)
- between 1 and 5 years (2)
- between 5 and 10 years (3)
- more than 10 years (4)
- there is no typical time frame (5)

B7 What is the main nature of research data used in your community?

- Quantitative (1)
- Qualitative (2)
- Both (3)

B8 What is the main type(s) of research data used in your community?

- Observations (1)
- Experiences (2)
- Simulations (3)
- Derived data (4)
- References (5)
- Digitizations (6)
- Excavations (7)
- Artefact or media artefact (9)
- Other: (8) _____

B8 + What is the typical size of datasets in your community?

- kB / ko (1)
- MB / Mo (2)
- GB / Go (3)
- TB / To (4)
- PB / Po (5)
- Other (6) _____
- I do not know (7)

ORD activities/practices

B9 Which answer seems to describe your skills at best in **producing a Data Management Plan (DMP)**?

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B9 + Which answer seems to describe your skills at best **in preparing your data for uploading them into a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B10 Which answer seems to describe your skills at best **in uploading research data into a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B11 Which answer seems to describe your skills at best **in sharing research data via a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Afficher cette question :

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

B12 Where does your community usually share its data?

- On a generalist repository (4)
- On an institutional repository (5)
- On a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Afficher cette question :

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

Et si

B12 != I do not know

B12 + Please indicate the name of the repository you use:

Afficher cette question :

If B11 = I am able to do it alone

Or B11 = I am able to do it with help from my community

Or B11 = I am able to do it with help from my institution

Or B11 = I am able to do it with help from the data infrastructure I use

Or B11 = I am able to do it with help from a librarian

B13 To which kind of data do you usually give access?

- Active data (4)
- Final data (preserved at the end of the project) (5)
- Both (6)

B14 Which answer seems to describe your skills at best **in searching research data produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

B15 Which answer seems to describe your skills at best **in re-using research data produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Afficher cette question :

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

B16 Which data do you re-use?

- Active data (1)
- Final data (preserved at the end of the project) (2)
- Both (3)

Afficher cette question :

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

B17 Where does your community usually get the data it re-uses?

- From a generalist repository (4)
- From an institutional repository (5)
- From a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Afficher cette question :

If B15 = I am able to do it alone

Or B15 = I am able to do it with help from my community

Or B15 = I am able to do it with help from my institution

Or B15 = I am able to do it with help from the data infrastructure I use

Or B15 = I am able to do it with help from a librarian

Et si

B17 != I do not know

B17 + Please indicate the name of the repository you use:

B18 Which answer seems to describe your skills at best **in reproducing research results produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from my librarian (6)
- I do not know if I am able to do it (7)

Standardization

B19 Do standards exist in your community?

- No (1)
- Not yet, but some are under development (2)
- Yes, but they are not widely used (3)
- Yes, they are widely used (4)
- I do not know (5)

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B20 Do the researchers of your community use **metadata standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B20 = Yes

B21 Which one(s)?

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B22 Do the researchers of your community use **PIDs standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B22 = Yes

B23 Which one(s)?

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B24 Do the researchers of your community use **file format standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B24 = Yes

B25 Which one(s)?

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B26 Do the researchers of your community use **documentation standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B26 = Yes

B27 Which one(s)?

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B28 Do the researchers of your community use **licence standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B28 = Yes

B29 Which one(s)?

Afficher cette question :

If B19 = Yes, they are widely used

Or B19 = Yes, but they are not widely used

B30 Do the researchers of your community use **other standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If B30 = Yes

B31 Which one(s)?

Obstacles/concerns/motivations

B32 What are **your practical obstacles** for sharing and reusing data? (please specify a value on the scale according to their importance)

Not important at all		Moderately important		Very important
1	2	3	4	5

There are restricted conditions on the use of the data that I handle (and thus on sharing them) ()	
I do not have enough time to prepare the data and to make them accessible and reusable ()	
The typical time scale of my research is very long (and I will not share my data before the project is finished) ()	
I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable ()	
I think that no appropriate technical tool is available for sharing my data and making them reusable ()	
I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable ()	
Other: ()	

Afficher cette question :

If B32 [There are restricted conditions on the use of the data that I handle (and thus on sharing them)] >= 2

B34 Which **restricted conditions** are there on the reuse of your data?

- legal restrictions (1)
- ethical restrictions (2)
- commercial restrictions (3)
- security restrictions (5)
- Other: (4) _____

B35 What are **your concerns** about following ORD practices? (please specify a value on the scale according to their importance)

Not important at all	Moderately important	Very important
1	2	3
4	5	

I am concerned about my data being not well handled and used ()	
I feel that I might lose control on my data ()	
I am concerned that someone else publishes results from my data before I do ()	
I am concerned not to be properly acknowledged as the data producer ()	
I think that the benefit would be too low as compared to the efforts ()	
I do not see the point ()	
Other: ()	

B37 What are **your motivations** to follow ORD practices? (please specify a value on the scale according to their importance)

	Not important at all	Moderately important	Very important		
	1	2	3	4	5
It complies with the requirements from funding agencies or publishers ()					
It brings me academic recognition (evaluation criterion for my academic career) ()					
It brings me scientific recognition ()					
I would like to benefit from the financial incentives existing for encouraging ORD practices ()					
It is very much practised in my community ()					
I commit to ORD because it corresponds to a personal belief ()					
Other: ()					

B39 Which actors could be more involved for better practices in ORD?

Fin de bloc: With (data) communities (A5 or A7 = YES)

Début de bloc: Without (data) communities (A5 and A7 = No/?)

Characterization of research

C1 Why?

- There is no community for my discipline (1)
- I mainly follow individual research practices (2)
- Other: (3) _____

C2 In a general way, to whom do you show your research data?

- Some colleagues in my lab (1)
- Some colleagues from the same research project or programme (2)
- Some colleagues from the same institution (regardless of discipline) (3)
- Some colleagues in the same discipline (4)
- General public (citizen science) (5)
- No one (6)
- Other: (8) _____
- I do not know (7)

C3 From whom the research data may be of interest to you?

- Some colleagues in my lab (1)
- Some colleagues from the same research project or programme (2)
- Some colleagues from the same institution (regardless of discipline) (3)
- Some colleagues in the same discipline (4)
- General public (citizen science) (5)
- No one (6)
- Other: (8) _____
- I do not know (7)

C5 What is the typical scale of your research?

- National (1)
- European (2)
- International (3)
- Other (4) _____

C6 What is the typical time frame of your research?

- less than 1 year (1)
- between 1 and 5 years (2)
- between 5 and 10 years (3)
- more than 10 years (4)
- there is no typical time frame (5)

C7 What is the main nature of your research data?

- Quantitative (1)
- Qualitative (2)
- Both (3)
- I do not use data (4)

Fin de bloc: Without (data) communities (A5 and A7 = No/?)

Début de bloc: Without (data) communities (C7 != 'I do not use data')

C8 What is the main type(s) of your research data?

- Observations (1)
- Experiments (2)
- Simulations (3)
- Derived data (4)
- References (5)
- Digitizations (6)
- Excavations (7)
- Artefact or media artefact (9)
- Other: (8) _____

C8 + What is the typical size of your datasets?

- kB / ko (1)
- MB / Mo (2)
- GB / Go (3)
- TB / To (4)
- PB / Po (5)
- Other: (6) _____
- I do not know (7)

ORD activities/practices

C9 Which answer seems to describe your skills at best **in producing a DMP?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C9 + Which answer seems to describe your skills at best **in preparing your data for uploading them into a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C10 Which answer seems to describe your skills at best **in uploading research data into a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C11 Which answer seems to describe your skills at best **in sharing research data via a repository?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Afficher cette question :

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

C12 Where do you usually share your data?

- On a generalist repository (4)
- On an institutional repository (5)
- On a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Afficher cette question :

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

Et si

C12 != I do not know

C12 + Please indicate the name of the repository you use:

Afficher cette question :

If C11 = I am able to do it alone

Or C11 = I am able to do it with help from my community

Or C11 = I am able to do it with help from my institution

Or C11 = I am able to do it with help from the data infrastructure I use

Or C11 = I am able to do it with help from a librarian

C13 To which kind of data do you usually give access?

- Active data (4)
- Final data (preserved at the end of the project) (5)
- Both (6)

C14 Which answer seems to describe your skills at best **in searching research data produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

C15 Which answer seems to describe your skills at best **in re-using research data produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Afficher cette question :

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

C16 Which data do you re-use?

- Active data (1)
- Final data (preserved at the end of the project) (2)
- Both (3)

Afficher cette question :

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

C17 Where do usually you get the data you re-use?

- From a generalist repository (4)
- From an institutional repository (5)
- From a disciplinary or specialized repository (6)
- Other: (8) _____
- I do not know (7)

Afficher cette question :

If C15 = I am able to do it alone

Or C15 = I am able to do it with help from my community

Or C15 = I am able to do it with help from my institution

Or C15 = I am able to do it with help from the data infrastructure I use

Or C15 = I am able to do it with help from a librarian

Et si

C17 != I do not know

C17 + Please indicate the name of the repository you use:

C18 Which answer seems to describe your skills at best **in reproducing research results produced by others?**

- I am aware of the process but I do not know how to do it (1)
- I am able to do it alone (2)
- I am able to do it with help from my community (3)
- I am able to do it with help from my institution (4)
- I am able to do it with help from the data infrastructure I use (5)
- I am able to do it with help from a librarian (6)
- I do not know if I am able to do it (7)

Standardization

C19 Are you aware of data standards?

- No (1)
- Not really, but I know that some of them are under development in my field (2)
- Yes, but they are not widely used in my field (3)
- Yes, they are widely used in my field (4)
- I do not know (5)

Afficher cette question :

If C19 = Yes, they are widely used in my field

Or C19 = Yes, but they are not widely used in my field

C20 Do you use **metadata standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C20 = Yes

C21 Which one(s)?

Afficher cette question :

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C22 Do you use **PIDs standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C22 = Yes

C23 Which one(s)?

Afficher cette question :

If C19 = Yes, they are widely used in my field

Or C19 = Yes, but they are not widely used in my field

C24 Do you use **file format standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C24 = Yes

C25 Which one(s)?

Afficher cette question :

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C26 Do you use **documentation standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C26 = Yes

C27 Which one(s)?

Afficher cette question :

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C28 Do you use **license standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C28 = Yes

C29 Which one(s)?

Afficher cette question :

If C19 = Yes, but they are not widely used in my field

Or C19 = Yes, they are widely used in my field

C30 Do you use **other standards**?

- No (2)
- Yes (1)
- I do not know (3)

Afficher cette question :

If C30 = Yes

C31 Which one(s)?

Obstacles/concerns/motivations

C32 What are your **practical obstacles** for sharing and reusing data? (please specify a value on the scale according to their importance)

	Not important at all	Moderately important	Very important		
	1	2	3	4	5
There are restricted conditions on the use of the data that I handle (and thus on sharing them) ()					
I do not have enough time to prepare the data and make them accessible and reusable ()					
The typical scale time of my research is very long (and I will not share my data before the project is finished) ()					
I feel that I do not have the right skills to prepare my data in a way that makes them accessible and reusable ()					
I think that no appropriate technical tool is available for sharing my data and making them reusable ()					
I do not have access to the necessary support or coaching that I would need to make my data accessible and reusable ()					
Other: ()					

Afficher cette question :

If C32 [There are restricted conditions on the use of the data that I handle (and thus on sharing them)] >= 2

C34 Which **restricted conditions** are there on the reuse of your data?

- legal restrictions (1)
- ethical restrictions (2)
- commercial restrictions (3)
- security restrictions (5)
- Other: (4) _____

C35 What are **your concerns** about following ORD practices? (please specify for them a value on the scale according to their importance)

	Not important at all	Moderately important	Very important		
	1	2	3	4	5
I am concerned about my data being not well handled and used ()					
I feel that I might lose control on my data ()					
I am concerned that someone else publishes results from my data before I do ()					
I am concerned not to be properly acknowledged as the data producer ()					
I think that the benefit would be too low as compared to the efforts ()					
I do not see the interest ()					
Other ()					

C37 What are **your motivations** to follow ORD practices? (please specify a value on the scale according to their importance)

Not important at all Moderately important Very important

1 2 3 4 5

It complies with the requirements from funding agencies or publishers ()	
It brings me academic recognition (evaluation criterion for my academic career) ()	
It brings me scientific recognition ()	
I would like to benefit from the financial incentives existing for encouraging ORD practices ()	
It is very much practised in my community ()	
I commit to ORD because it corresponds to a personal belief ()	
Other: ()	

C39 Which actors could be more involved for better practices in ORD?

Fin de bloc: Without (data) communities (C7 != 'I do not use data')

Début de bloc: Without (data) communities (C7 = 'I do not use data')

D1 What are the main particularities of your research?

D4 Who are the main funders of your research?

D2 Do you publish in Open Access?

- Never (3)
- Sometimes (2)
- Frequently (1)
- I am not sure about what it means (4)

Annexe 5 : Liste des sites web issus des réponses à l'enquête

ala.org	https://www.euro-online.org/websites/bor/
aforngen.wsl.ch/	https://www.europeansociology.org/
atlas.cern	https://www.gcb.uzh.ch/en.html
dada.zahnd.be	https://www.geo.uzh.ch/en/units/rsi
dhd-blog.org	https://www.globalbuddhism.org/jgb/index.php/jgb/
dkf.unibas.ch	https://www.hes-so.ch/accueil
exoplanets.eu	https://www.hets-fr.ch/fr/recherche/accueil
github.com	https://www.hetsl.ch/laress/
hfh.ch/sure	https://www.hevs.ch/fr/rad-instituts/institut-sante/
historikerinnen.ch	https://www.hevs.ch/fr/rad-instituts/institut-sante/activites-instituts/physiolab-11160
hssuisse.ch	https://www.iassidd.org/
http://aclals.net/	https://www.iavceivolcano.org
http://arthobservations.org	https://www.isanet.org
http://contemporarydrama.de/	https://www.isle-linguistics.org/
http://enfin.info/	https://www.linkedin.com/groups/13763674/
http://p3.snf.ch/Project-180350	https://www.manep.ch/
http://swsa.semanticweb.org	https://www.monitoringdemocracy.eu/
http://www.airhm.net/delegations/delegation-suisse/	https://www.nccr-catalysis.ch/
http://www.comparativecandidates.org/	https://www.postcolonialstudiesassociation.co.uk/
http://www.fremdsprachendidaktik.org/	https://www.sagw.ch/fr/seg/,
http://www.irscl.com/	https://www.saute.ch
https://aaac.world/	https://www.saute.ch/en/saute/
https://aom.org	https://www.sgg-ssl.ch/sgg/
https://blog.unifr.ch/carbonatelab/	https://www.sgr-sssr.ch/sgr-sssr/
https://brain.ieee.org	https://www.sgs-sss.ch/
https://cigev.unige.ch/	https://www.sgs-sss.ch/en/
https://citizensciences.net/	https://www.sib.swiss/
https://claire-ai.org	https://www.spectralbiology.org/
https://conbio.org	https://www.springer.com/journal/146
https://cses.org/	https://www.uni-due.de/biwi/politische-bildung/forschungskolloquium.php
https://data-innovation.org/	https://www.unifr.ch/cgf/en/research/research.html
https://dh-ch.ch/	https://www.unifr.ch/ecopol/fr/
https://dh-ch.ch/	https://www.unifr.ch/english/en/projects/civility-project/
https://dig-hum.de/	https://www.unifr.ch/med/de/research/groups/nordmann/
https://digitale-kunstgeschichte.de/	https://www.unige.ch/fapse/decisionlab/
https://digitalpsychotherapylab.ch	https://www.unige.ch/sciences/astro/variability/en/
https://droit.cuso.ch/accueil	https://www.unige.ch/sciences-societe/socio/fr/recherche/irs/index/
https://eacials.eu/	https://www.unil.ch/cemep/fr/home.html
https://essenglish.org/	https://www.unil.ch/samemes/en/home.html

https://forum.openhardware.science/	https://www.unilu.ch/fakultaeten/tf/
https://gleon.org/	https://www.wsl.ch/de/ueber-die-wsl/programme-und-initiativen/forschungsinitiative-blue-green-biodiversity.html
https://hia4sd.net/	https://www.wsl.ch/en/about-wsl/programmes-and-initiatives/wsl-biodiversity-center.html
https://hummingbird.bio/	https://dasch.swiss
https://ibam.swiss/fr/home-2-main-fr/	infoclio.ch
https://iiif.io	many. Ei. infoclio.ch
https://institut-mehrsprachigkeit.ch/deit	nccr-onthemove.ch/
https://linked.art/	safmed.ch
https://lives-nccr.ch	scala-lang.org
https://mitglied.scnat.ch/sgw-ssbf	scta.info
https://musewiki.wikidot.com/museestmeetings	sites.google.com/fmach.it/g-bike-genetics-eu/home
https://nccr-onthemove.ch/	sjdm.org
https://nccr-robotics.ch/	stackexchange.com
https://pragmatics.international/default.aspx	stackoverflow.com
https://research.cs.wisc.edu/dbworld/browse.html	unige.ch/adp
https://researchdata.unibas.ch/en/	www.eahil.eu
https://sigchi.org/	www.ecargument.org
https://sslarch.github.io	www.esag.swiss
https://swissbias.github.io/	www.evoltree.eu
https://swissforestlab.wsl.ch/de/index.html	www.federalism.ch
https://unfccc.int/topics/land-use/workstreams/land-use-land-use-change-and-forestry-lulucf	www.gentree-h2020.eu
https://www.4sonline.org/	www.ighg.org
https://www.aric-interculturel.com/	www.lavater.com
https://www.asianstudies.org/	www.mlanet.org
https://www.aslo.org/	www.ophen.org
https://www.bioinspired-materials.ch/en/	www.pancare.eu
https://www.centre-lives.ch/	www.pancarefollowup.eu
https://www.dgff.de/	www.publicpolicyargument.eu
https://www.earli.org	www.Rai.uk
https://www.earlychildhoodresearch.ch/	www.selects.ch
https://www.eawag.ch/en/research/water-for-ecosystem/biodiversity/blue-green-biodiversity-research-initiative/	www.sphn.ch
https://www.ecolelasource.ch/vieillissement-sante/	www.spog.ch
https://www.egu.eu/cr/home/	www.zne.uzh.ch
https://www.elmi2021.org/	

Annexe 6 : Liste de sites ou pages Internet rattachées à des communautés de données

Selon les réponses des chercheurs données dans l'enquête

https://dh-ch.ch/
https://musewiki.wikidot.com/musestmeetings
https://dig-hum.de/
https://nccr-robotics.ch/
https://www.sib.swiss/
http://arthobservations.org
http://enfin.info/
https://www.gcb.uzh.ch/en.html
https://www.elmi2021.org/
https://hummingbird.bio/
https://iiif.io/
www.sphn.ch
atlas.cern
https://www.earlychildhoodresearch.ch/
https://www.aric-interculturel.com/
http://www.comparativecandidates.org/
https://www.unige.ch/urbanhub/bienvenue/
https://forum.openhardware.science/
infoclio.ch
www.esag.swiss
https://unfccc.int/topics/land-use/workstreams/land-use--land-use-change-and-forestry-lulucf
https://cses.org/
https://www.monitoringdemocracy.eu/
https://www.geo.uzh.ch/en/units/rsl
https://www.spectralbiology.org/
https://swissforestlab.wsl.ch/de/index.html